Third Edition

ABSTRACT ALGEBRA

I. N. Herstein

2

GROUPS

https://www.facebook.com/GrinchaZ

1. DEFINITIONS AND EXAMPLES OF GROUPS

We have seen in Section 4 of Chapter 1 that given any nonempty set, the set $A(S)$ of all 1-1 mappings of S onto itself is not just a set alone, but has a far richer texture. The possibility of combining two elements of $A(S)$ to get yet another element of $A(S)$ endows $A(S)$ with an algebraic structure. We recall how this was done: If $f, g \in A(S)$, then we combine them to form the mapping fg defined by $(fg)(s) = f(g(s))$ for every $s \in S$. We called fg the *product* of f and g, and verified that $fg \in A(S)$, and that this product obeyed certain rules. From the myriad of possibilities we somehow selected four particular rules that govern the behavior of $A(S)$ relative to this product.

These four rules were

1. *Closure*, namely if $f, g \in A(S)$, then $fg \in A(S)$. We say that $A(S)$ is *closed* under this product.
2. *Associativity*, that is, given $f, g, h \in A(S)$, then $f(gh) = (fg)h$.
3. *Existence of a unit element*, namely, there exists a particular element $i \in A(S)$ (the identity mapping) such that $fi = if = f$ for all $f \in A(S)$.
4. *Existence of inverses*, that is, given $f \in A(S)$ there exists an element, denoted by f^{-1}, in $A(S)$ such that $ff^{-1} = f^{-1}f = i$.

To justify or motivate why these four specific attributes of $A(S)$ were singled out, in contradistinction to some other set of properties, is not easy to

do. In fact, in the history of the subject it took quite some time to recognize that these four properties played the key role. We have the advantage of historical hindsight, and with this hindsight we choose them not only to study $A(S)$, but also as the chief guidelines for abstracting to a much wider context.

Although we saw that the four properties above enabled us to calculate concretely in $A(S)$, there were some differences with the kind of calculations we are used to. If S has three or more elements, we saw in Problem 15, Chapter 1, Section 4 that it is possible for $f, g \in A(S)$ to have $fg \neq gf$. However, this did not present us with insurmountable difficulties.

Without any further polemics we go to the

Definition. A nonempty set G is said to be a *group* if in G there is defined an operation $*$ such that:

(a) $a, b \in G$ implies that $a * b \in G$. (We describe this by saying that G is *closed* under $*$.)

(b) Given $a, b, c \in G$, then $a * (b * c) = (a * b) * c$. (This is described by saying that the *associative law* holds in G.)

(c) There exists a special element $e \in G$ such that $a * e = e * a = a$ for all $a \in G$ (e is called the *identity* or *unit element* of G).

(d) For every $a \in G$ there exists an element $b \in G$ such that $a * b = b * a = e$. (We write this element b as a^{-1} and call it the *inverse* of a in G.)

These four defining postulates (called the *group axioms*) for a group were, after all, patterned after those that hold in $A(S)$. So it is not surprising that $A(S)$ is a group relative to the operation "composition of mappings."

The operation $*$ in G is usually called the *product*, but keep in mind that this has nothing to do with product as we know it for the integers, rationals, reals, or complexes. In fact, as we shall see below, in many familiar examples of groups that come from numbers, what we call the product in these groups is actually the addition of numbers. *However, a general group need have no relation whatsoever to a set of numbers.* We reiterate: A group is no more, no less, than a nonempty set with an operation $*$ satisfying the four group axioms.

Before starting to look into the nature of groups, we look at some examples.

Examples of Groups

1. Let $\mathbb{Z}$ be the set of all integers and let $*$ be the ordinary addition, $+$, in $\mathbb{Z}$. That $\mathbb{Z}$ is closed and associative under $*$ are basic properties of the integers. What serves as the unit element, e, of $\mathbb{Z}$ under $*$? Clearly, since $a = a * e =$

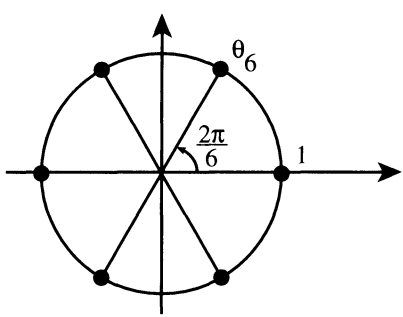
$a + e$, we have $e = 0$, and 0 is the required identity element under addition. What about a^{-1}? Here too, since $e = 0 = a * a^{-1} = a + a^{-1}$, the a^{-1} in this instance is $-a$, and clearly $a * (-a) = a + (-a) = 0$.

2. Let $\mathbb{Q}$ be the set of all rational numbers and let the operation $*$ on $\mathbb{Q}$ be the ordinary addition of rational numbers. As above, $\mathbb{Q}$ is easily shown to be a group under $*$. Note that $\mathbb{Z} \subset \mathbb{Q}$ and both $\mathbb{Z}$ and $\mathbb{Q}$ are groups under the same operation $*$.

3. Let $\mathbb{Q}'$ be the set of all *nonzero* rational numbers and let the operation $*$ on $\mathbb{Q}'$ be the ordinary multiplication of rational numbers. By the familiar properties of the rational numbers we see that $\mathbb{Q}'$ forms a group relative to $*$.

4. Let $\mathbb{R}^+$ be the set of all *positive real* numbers and let the operation $*$ on $\mathbb{R}^+$ be the ordinary product of real numbers. Again it is easy to check that $\mathbb{R}^+$ is a group under $*$.

5. Let E_n be the set of θ_n^i, $i = 0, 1, 2, \ldots, n-1$, where θ_n is the complex number $\theta_n = \cos(2\pi/n) + i \sin(2\pi/n)$. Let $\theta_n^k * \theta_n^j = \theta_n^{k+j}$, the ordinary product of the powers of θ_n as complex numbers. By De Moivre's Theorem we saw that $\theta_n^n = 1$. We leave it to the reader to verify that E_n is a group under $*$. The elements of E_n are called the *n'th roots of unity.* The picture below illustrates the group E_6, whose elements are represented by the dots on the unit circle in the complex plane.



Note one striking difference between the Examples 1 to 4 and Example 5; the first four have an infinite number of elements, whereas E_n has a finite number, n, of elements.

Definition. A group G is said to be a *finite group* if it has a finite number of elements. The number of elements in G is called the *order* of G and is denoted by $|G|$.

Thus E_n above is a finite group, and $|E_n| = n$.

All the examples presented above satisfy the additional property that $a * b = b * a$ for any pair of elements. This need not be true in a group. Just witness the case of $A(S)$, where S has three or more elements; there we saw that we could find $f, g \in A(S)$ such that $fg \neq gf$.

This prompts us to single out as special those groups of G in which $a * b = b * a$ for all $a, b \in G$.

Definition. A group G is said to be *abelian* if $a * b = b * a$ for all $a, b \in G$.

The word abelian derives from the name of the great Norwegian mathematician Niels Henrik Abel (1802–1829), one of the greatest scientists Norway has ever produced.

A group that is not abelian is called *nonabelian*, a not too surprising choice of name.

We now give examples of some nonabelian groups. Of course, the $A(S)$ afford us an infinite family of such. But we present a few other examples in which we can compute quite readily.

6. Let $\mathbb{R}$ be the set of all real numbers, and let G be the set of all mappings $T_{a,b}: \mathbb{R} \to \mathbb{R}$ defined by $T_{a,b}(r) = ar + b$ for any real number r, where a, b are real numbers and $a \neq 0$. Thus, for instance, $T_{5,-6}$ is such that $T_{5,-6}(r) = 5r - 6$; $T_{5,-6}(14) = 5 \cdot 14 - 6 = 64$, $T_{5,-6}(\pi) = 5\pi - 6$. The $T_{a,b}$ are 1-1 mappings of $\mathbb{R}$ onto itself, and we let $T_{a,b} * T_{c,d}$ be the product of two of these mappings. So

$$(T_{a,b} * T_{c,d})(r) = T_{a,b}(T_{c,d}(r)) = aT_{c,d}(r) + b = a(cr + d) + b$$
$$= (ac)r + (ad + b) = T_{ac,\,ad+b}(r).$$

So we have the formula

$$T_{a,b} * T_{c,d} = T_{ac,\,ad+b}. \tag{1}$$

This result shows us that $T_{a,b} * T_{c,d}$ is in G—for it satisfies the membership requirement for belonging to G—so G is closed under $*$. Since we are talking about the product of mappings (i.e., the composition of mappings), $*$ is associative. The element $T_{1,0} = i$ is the identity mapping of $\mathbb{R}$ onto itself. Finally, what is $T_{a,b}^{-1}$? Can we find real numbers $x \neq 0$ and y, such that

$$T_{a,b} * T_{x,y} = T_{x,y} * T_{a,b} = T_{1,0}?$$

Go back to (1) above; we thus want $T_{ax,\,ay+b} = T_{1,0}$, that is, $ax = 1$, $ay + b = 0$. Remember now that $a \neq 0$, so if we put $x = a^{-1}$ and $y = -a^{-1}b$, the required relations are satisfied. One verifies immediately that

$$T_{a,b} * T_{a^{-1},-a^{-1}b} = T_{a^{-1},-a^{-1}b} * T_{a,b} = T_{1,0}.$$

So G is indeed a group.

What is $T_{c,d} * T_{a,b}$? According to the formula given in (1), where we replace a by c, c by a, b by d, d by b, we get

$$T_{c,d} * T_{a,b} = T_{ca,\,cb+d}. \tag{2}$$

Thus $T_{c,d} * T_{a,b} =$ if $T_{a,b} * T_{c,d}$ and only if $bc + d = ad + b$. This fails to be true, for instance, if $a = 1$, $b = 1$, $c = 2$, $d = 3$. So G is nonabelian.

7. Let $H \subset G$, where G is the group in Example 6, and H is defined by $H = \{T_{a,b} \in G \mid a \text{ is rational, } b \text{ any real}\}$. We leave it to the reader to verify that H is a group under the operation $*$ defined on G. H is nonabelian.

8. Let $K \subset H \subset G$, where H, G are as above and $K = \{T_{1,b} \in G \mid b \text{ any real}\}$. The reader should check that K is a group relative to the operation $*$ of G, and *that K is, however, abelian.*

9. Let S be the plane, that is, $S = \{(x, y) \mid x, y \text{ real}\}$ and consider $f, g \in A(S)$ defined by $f(x, y) = (-x, y)$ and $g(x, y) = (-y, x)$; f is the reflection about the y-axis and g is the rotation through 90° in a counterclockwise direction about the origin. We then define $G = \{f^i g^j \mid i = 0, 1; j = 0, 1, 2, 3\}$, and let $*$ in G be the product of elements in $A(S)$. Clearly, $f^2 = g^4 =$ identity mapping;

$$(f * g)(x, y) = (fg)(x, y) = f(g(x, y)) = f(-y, x) = (y, x)$$

and

$$(g * f)(x, y) = g(f(x, y)) = g(-x, y) = (-y, -x).$$

So $g * f \neq f * g$. We leave it to the reader to verify that $g * f = f * g^{-1}$ and G is a nonabelian group of order 8. This group is called the *dihedral group* of order 8. [Try to find a formula for $(f^i g^j) * (f^s g^t) = f^a g^b$ that expresses a, b in terms of i, j, s, and t.]

10. Let S be as in Example 9 and f the mapping in Example 9. Let $n > 2$ and let h be the rotation of the plane about the origin through an angle of $2\pi/n$ in the counterclockwise direction. We then define $G = \{f^k h^j \mid k = 0, 1; j = 0, 1, 2, \ldots, n - 1\}$ and define the product $*$ in G via the usual product of mappings. One can verify that $f^2 = h^n =$ identity mapping, and $fh = h^{-1}f$.

These relations allow us to show (with some effort) that G is a nonabelian group of order $2n$. G is called the *dihedral group* of order $2n$.

11. Let $G = \{ f \in A(S) \mid f(s) \neq s$ for only a finite number of $s \in S\}$, where we suppose that S is an *infinite* set. We claim that G is a group under the product $*$ in $A(S)$. The associativity holds automatically in G, since it already holds in $A(S)$. Also, $i \in G$, since $i(s) = s$ for *all* $s \in S$. So we must show that G is closed under the product and if $f \in G$, then $f^{-1} \in G$.

We first dispose of the closure. Suppose that $f, g \in G$; then $f(s) = s$ except, say, for $s_1, s_2, \ldots, s_n$ and $g(s) = s$ except for $s_1', s_2', \ldots, s_m'$. Then $(fg)(s) = f(g(s)) = s$ for all s other than $s_1, s_2, \ldots, s_n, s_1', \ldots, s_m'$ (and possibly even for some of these). So fg moves only a finite number of elements of S, so $fg \in G$.

Finally, if $f(s) = s$ for all s other than $s_1, s_2, \ldots, s_n$, then $f^{-1}(f(s)) = f^{-1}(s)$, but $f^{-1}(s) = f^{-1}(f(s)) = (f^{-1}f)(s) = i(s) = s$. So we obtain that $f^{-1}(s) = s$ for all s except $s_1, \ldots, s_n$. Thus $f^{-1} \in G$ and G satisfies all the group axioms, hence G is a group.

12. Let G be the set of all mappings T_θ, where T_θ is the rotation of a given circle about its center through an angle θ in the clockwise direction. In G define $*$ by the composition of mappings. Since, as is readily verified, $T_\theta * T_\psi = T_{\theta+\psi}$, G is closed under $*$. The other group axioms check out easily. Note that $T_{2\pi} = T_0 =$ the identity mapping, and $T_\theta^{-1} = T_{-\theta} = T_{2\pi-\theta}$. G is an abelian group.

As we did for $A(S)$ we introduce the shorthand notation a^n for

$$\underbrace{a * a * a \cdots * a}_{n \text{ times}}$$

and define $a^{-n} = (a^{-1})^n$, for n a positive integer, and $a^0 = e$. The usual rules of exponents then hold, that is, $(a^m)^n = a^{mn}$ and $a^m * a^n = a^{m+n}$ for *any* integers m and n.

Note that with this notation, if G is the group of integers under $+$, then a^n is really na.

Having seen the 12 examples of groups above, the reader might get the impression that all, or almost all, sets with some operation $*$ form groups. This is far from true. We now give some examples of nongroups. In each case we check the four group axioms and see which of these fail to hold.

Nonexamples

1. Let G be the set of all integers, and let $*$ be the ordinary product of integers in G. Since $a * b = ab$, for $a, b \in G$ we clearly have that G is closed and associative relative to $*$. Furthermore, the number 1 serves as the unit ele-

ment, since $a * 1 = a1 = a = 1a = 1 * a$ for every $a \in G$. So we are three-fourths of the way to proving that G is a group. All we need is inverses for the elements of G, relative to $*$, to lie in G. But this just isn't so. Clearly, we cannot find an integer b such that $0 * b = 0b = 1$, since $0b = 0$ for all b. But even other integers fail to have inverses in G. For instance, we *cannot find an integer* b such that $3 * b = 1$ (for this would require that $b = \frac{1}{3}$, and $\frac{1}{3}$ is not an integer).

2. Let G be the set of all nonzero real numbers and define, for $a, b \in G$, $a * b = a^2b$; thus $4 * 5 = 4^2(5) = 80$. Which of the group axioms hold in G under this operation $*$ and which fail to hold? Certainly, G is closed under $*$. Is $*$ associative? If so, $(a * b) * c = a * (b * c)$, that is, $(a * b)^2c = a^2(b * c)$, and so $(a^2b)^2c = a^2(b^2c)$, which boils down to $a^2 = 1$, which holds only for $a = \pm 1$. So, in general, the associative law does *not* hold in G relative to $*$. We similarly can verify that G does not have a unit element. Thus even to discuss inverses relative to $*$ would not make sense.

3. Let G be the set of all *positive* integers, under $*$ where $a * b = ab$, the ordinary product of integers. Then one can easily verify that G fails to be a group *only because it fails* to have inverses for some (in fact, most) of its elements relative to $*$.

We shall find some other nonexamples of groups in the exercises.

PROBLEMS

Easier Problems

1. Determine if the following sets G with the operation indicated form a group. If not, point out which of the group axioms fail.
- **(a)** G = set of all integers, $a * b = a - b$.
- **(b)** G = set of all integers, $a * b = a + b + ab$.
- **(c)** G = set of nonnegative integers, $a * b = a + b$.
- **(d)** G = set of all rational numbers $\neq -1$, $a * b = a + b + ab$.
- **(e)** G = set of all rational numbers with denominator divisible by 5 (written so that numerator and denominator are relatively prime), $a * b = a + b$.
- **(f)** G a set having more than one element, $a * b = a$ for all $a, b \in G$.

2. In the group G defined in Example 6, show that the set $H = \{T_{a,b} \mid a = \pm 1, b \text{ any real}\}$ forms a group under the $*$ of G.

3. Verify that Example 7 is indeed an example of a group.

4. Prove that K defined in Example 8 is an abelian group.
5. In Example 9, prove that $g * f = f * g^{-1}$, and that G is a group, is nonabelian, and is of order 8.
6. Let G and H be as in Examples 6 and 7, respectively. Show that if $T_{a,b} \in G$, then $T_{a,b} * V * T_{a,b}^{-1} \in H$ if $V \in H$.
7. Do Problem 6 with H replaced by the group K of Example 8.
8. If G is an abelian group, prove that $(a * b)^n = a^n * b^n$ for all integers n.
9. If G is a group in which $a^2 = e$ for all $a \in G$, show that G is abelian.
10. If G is the group in Example 6, find all $T_{a,b} \in G$ such that $T_{a,b} * T_{1,x} = T_{1,x} * T_{a,b}$ for *all* real x.
11. In Example 10, for $n = 3$ find a formula that expresses $(f^i h^j) * (f^s h^t)$ as $f^a h^b$. Show that G is a nonabelian group of order 6.
12. Do Problem 11 for $n = 4$.
13. Show that any group of order 4 or less is abelian.
14. If G is any group and $a, b, c \in G$, show that if $a * b = a * c$, then $b = c$, and if $b * a = c * a$, then $b = c$.
15. Express $(a * b)^{-1}$ in terms of a^{-1} and b^{-1}.
16. Using the result of Problem 15, prove that a group G in which $a = a^{-1}$ for every $a \in G$ must be abelian.
17. In any group G, prove that $(a^{-1})^{-1} = a$ for all $a \in G$.
18. If G is a finite group of *even* order, show that there must be an element $a \neq e$ such that $a = a^{-1}$. (**Hint:** Try to use the result of Problem 17.)
19. In S_3, show that there are four elements x satisfying $x^2 = e$ and three elements y satisfying $y^3 = e$.
20. Find all the elements in S_4 such that $x^4 = e$.

Middle-Level Problems

21. Show that a group of order 5 must be abelian.
22. Show that the set defined in Example 10 is a group, is nonabelian, and has order $2n$. Do this by finding the formula for $(f^i h^j) * (f^s h^t)$ in the form $f^a h^b$.
23. In the group G of Example 6, find all elements $U \in G$ such that $U * T_{a,b} = T_{a,b} * U$ for every $T_{a,b} \in G$.
24. If G is the dihedral group of order $2n$ as defined in Example 10, prove that:
    (a) If n is odd and $a \in G$ is such that $a * b = b * a$ for all $b \in G$, then $a = e$.
    (b) If n is even, show that there is an $a \in G$, $a \neq e$, such that $a * b = b * a$ for all $b \in G$.

(c) If n is even, find all the elements $a \in G$ such that $a * b = b * a$ for all $b \in G$.

25. If G is any group, show that:

(a) e is unique (i.e., if $f \in G$ also acts as a unit element for G, then $f = e$).

(b) Given $a \in G$, then $a^{-1} \in G$ is unique.

***26.** If G is a finite group, prove that, given $a \in G$, there is a positive integer n, depending on a, such that $a^n = e$.

***27.** In Problem 26, show that there is an integer $m > 0$ such that $a^m = e$ for *all* $a \in G$.

Harder Problems

28. Let G be a set with an operation $*$ such that:

1. G is closed under $*$.
2. $*$ is associative.
3. There exists an element $e \in G$ such that $e * x = x$ for all $x \in G$.
4. Given $x \in G$, there exists a $y \in G$ such that $y * x = e$.

Prove that G is a group. (Thus you must show that $x * e = x$ and $x * y = e$ for e, y as above.)

29. Let G be a *finite* nonempty set with an operation $*$ such that:

1. G is closed under $*$.
2. $*$ is associative.
3. Given $a, b, c \in G$ with $a * b = a * c$, then $b = c$.
4. Given $a, b, c, \in G$ with $b * a = c * a$, then $b = c$.

Prove that G must be a group under $*$.

30. Give an example to show that the result of Problem 29 can be false if G is an infinite set.

31. Let G be the group of all nonzero real numbers under the operation $*$ which is the ordinary multiplication of real numbers, and let H be the group of all real numbers under the operation #, which is the addition of real numbers.

(a) Show that there is a mapping $F: G \to H$ of G onto H which satisfies $F(a * b) = F(a) \# F(b)$ for all $a, b \in G$ [i.e., $F(ab) = F(a) + F(b)$].

(b) Show that no such mapping F can be 1-1.

2. SOME SIMPLE REMARKS

In this short section we show that certain formal properties which follow from the group axioms hold in any group. As a matter of fact, most of these results have already occurred as problems at the end of the preceding section.

It is a little clumsy to keep writing the $*$ for the product in G, and *from now on we shall write the product $a * b$ simply as ab* for all $a, b \in G$.

The first such formal results we prove are contained in

Lemma 2.2.1. If G is a group, then:

(a) Its identity element is *unique*.
(b) Every $a \in G$ has a *unique* inverse $a^{-1} \in G$.
(c) If $a \in G$, $(a^{-1})^{-1} = a$.
(d) For $a, b \in G$, $(ab)^{-1} = b^{-1}a^{-1}$.

Proof. We start with Part (a). What is expected of us to carry out the proof? We must show that if $e, f \in G$ and $af = fa = a$ for all $a \in G$ and $ae = ea = a$ for all $a \in G$, then $e = f$. This is very easy, for then $e = ef$ and $f = ef$; hence $e = ef = f$, as required.

Instead of proving Part (b), we shall prove a stronger result (listed below as Lemma 2.2.2), which will have Part (b) as an immediate consequence. We claim that in a group G if $ab = ac$, then $b = c$; that is, we can *cancel a given element from the same side of an equation*. To see this, we have, for $a \in G$, an element $u \in G$ such that $ua = e$. Thus from $ab = ac$ we have

$$u(ab) = u(ac),$$

so, by the associative law, $(ua)b = (ua)c$, that is, $eb = ec$. Hence $b = eb = ec = c$, and our result is established. A similar argument shows that if $ba = ca$, then $b = c$. However, we *cannot conclude* from $ab = ca$ that $b = c$; in any abelian group, yes, but in general, no.

Now to get Part (b) as an implication of the cancellation result. Suppose that $b, c \in G$ act as inverses for a; then $ab = e = ac$, so by cancellation $b = c$ and we see that the inverse of a is unique. We shall always write it as a^{-1}.

To see Part (c), note that by definition $a^{-1}(a^{-1})^{-1} = e$; but $a^{-1}a = e$, so by cancellation in $a^{-1}(a^{-1})^{-1} = e = a^{-1}a$ we get that $(a^{-1})^{-1} = a$.

Finally, for Part (d) we calculate

$$\begin{aligned}(ab)(b^{-1}a^{-1}) &= ((ab)b^{-1})a^{-1} \qquad \text{(associative law)} \\ &= (a(bb^{-1})a^{-1} \qquad \text{(again the associative law)} \\ &= (ae)a^{-1} = aa^{-1} = e.\end{aligned}$$

Similarly, $(b^{-1}a^{-1})(ab) = e$. Hence, by definition, $(ab)^{-1} = b^{-1}a^{-1}$. □

We promised to list a piece of the argument given above as a separate lemma. We keep this promise and write

Lemma 2.2.2. In any group G and $a, b, c \in G$, we have:

(a) If $ab = ac$, then $b = c$.

(b) If $ba = ca$, then $b = c$.

Before leaving these results, note that if G is the group of real numbers under $+$, then Part (c) of Lemma 2.2.1 translates into the familiar $-(-a) = a$.

There is only a scant bit of mathematics in this section; accordingly, we give only a few problems. No indication is given as to the difficulty of these.

PROBLEMS

1. Suppose that G is a set closed under an associative operation such that
1. given $a, y \in G$, there is an $x \in G$ such that $ax = y$, and
2. given $a, w \in G$, there is a $u \in G$ such that $ua = w$.

Show that G is a group.

***2.** If G is a *finite* set closed under an associative operation such that $ax = ay$ forces $x = y$ and $ua = wa$ forces $u = w$, for every $a, x, y, u, w \in G$, prove that G is a group. (This is a repeat of a problem given earlier. It will be used in the body of the text later.)

3. If G is a group in which $(ab)^i = a^i b^i$ for three consecutive integers i, prove that G is abelian.

4. Show that the result of Problem 3 would not always be true if the word "three" were replaced by "two." In other words, show that there is a group G and consecutive numbers $i, i + 1$ such that G is not abelian but does have the property that $(ab)^i = a^i b^i$ and $(ab)^{i+1} = a^{i+1} b^{i+1}$ for all a, b in G.

5. Let G be a group in which $(ab)^3 = a^3 b^3$ and $(ab)^5 = a^5 b^5$ for all $a, b \in G$. Show that G is abelian.

6. Let G be a group in which $(ab)^n = a^n b^n$ for some fixed integer $n > 1$ for all $a, b \in G$. For all $a, b \in G$, prove that:

(a) $(ab)^{n-1} = b^{n-1} a^{n-1}$.

(b) $a^n b^{n-1} = b^{n-1} a^n$.

(c) $(aba^{-1}b^{-1})^{n(n-1)} = e$.

[Hint for Part (c): Note that $(aba^{-1})^r = ab^r a^{-1}$ for all integers r.]

3. SUBGROUPS

In order for us to find out more about the makeup of a given group G, it may be too much of a task to tackle all of G head-on. It might be desirable to focus our attention on appropriate pieces of G, which are smaller, over which we have some control, and are such that the information gathered about them can be used to get relevant information and insight about G itself. The question then becomes: What should serve as suitable pieces for this kind of dissection of G? Clearly, whatever we choose as such pieces, we want them to reflect the fact that G is a group, not merely any old set.

A group is distinguished from an ordinary set by the fact that it is endowed with a well-behaved operation. It is thus natural to demand that such pieces above behave reasonably with respect to the operation of G. Once this is granted, we are led almost immediately to the concept of a subgroup of a group.

Definition. A nonempty subset, H, of a group G is called a *subgroup* of G if, relative to the product in G, H itself forms a group.

We stress the phrase "relative to the product in G." Take, for instance, the subset $A = \{1, -1\}$ in $\mathbb{Z}$, the set of integers. Under the multiplication of integers, A is a group. But A is *not* a subgroup of $\mathbb{Z}$ viewed as a group with respect to $+$.

Every group G automatically has two obvious subgroups, namely G itself and the subgroup consisting of the identity element, e, alone. These two subgroups we call *trivial subgroups*. Our interest will be in the remaining ones, the *proper subgroups* of G.

Before proceeding to a closer look at the general character of subgroups, we want to look at some specific subgroups of some particular, explicit groups. Some of the groups we consider are those we introduced as examples in Section 1; we maintain the numbering given there for them. In some of these examples we shall verify that certain specified subsets are indeed subgroups. We would strongly recommend that the reader carry out such a verification in lots of the others and try to find other examples for themselves.

In trying to verify whether or not a given subset of a group is a subgroup, we are spared checking one of the axioms defining a group, namely the associative law. Since the associative law holds universally in a group G, given any subset A of G and any three elements of A, then the associative law certainly holds for them. So we must check, for a given subset A of G, whether A is closed under the operation of G, whether e is in A, and finally, given $a \in A$, whether a^{-1} is also in A.

Note that we can save one more calculation. Suppose that $A \subset G$ is nonempty and that given $a, b \in A$, then $ab \in A$. Suppose further that given $a \in A$, then $a^{-1} \in A$. Then we assert that $e \in A$. For pick $a \in A$; then $a^{-1} \in A$ by supposition, hence $aa^{-1} \in A$, again by supposition. Since $aa^{-1} = e$, we have that $e \in A$. Thus a is a subgroup of G. In other words,

Lemma 2.3.1. A nonempty subset $A \subset G$ is a subgroup of G if and only if A is closed with respect to the operation of G and, given $a \in A$, then $a^{-1} \in A$.

We now consider some examples.

Examples

1. Let G be the group $\mathbb{Z}$ of integers under $+$ and let H be the set of even integers. We claim that H is a subgroup of $\mathbb{Z}$. Why? Is H closed, that is, given $a, b \in H$, is $a + b \in H$? In other words, if a, b are even integers, is $a + b$ an even integer? The answer is yes, so H is certainly closed under $+$. Now to the inverse. Since the operation in $\mathbb{Z}$ is $+$, the inverse of $a \in \mathbb{Z}$ relative to this operation is $-a$. If $a \in H$, that is, if a is even, then $-a$ is also even, hence $-a \in H$. In short, H is a subgroup of $\mathbb{Z}$ under $+$.

2. Let G once again be the group $\mathbb{Z}$ of integers under $+$. In Example 1, H, the set of even integers, can be described in another way: namely H consists of all multiples of 2. There is nothing particular in Example 1 that makes use of 2 itself. Let $m > 1$ be any integer and let H_m consist of all multiples of m in $\mathbb{Z}$. We leave it to the reader to verify that H_m is a subgroup of $\mathbb{Z}$ under $+$.

3. Let S be any nonempty set and let $G = A(S)$. If $a \in S$, let $H(a) = \{f \in A(S) \mid f(a) = a\}$. We claim that $H(a)$ is a subgroup of G. For if $f, g \in H(a)$, then $(fg)(a) = f(g(a)) = f(a) = a$, since $f(a) = g(a) = a$. Thus $fg \in H(a)$. Also, if $f \in H(a)$, then $f(a) = a$, so that $f^{-1}(f(a)) = f^{-1}(a)$. But $f^{-1}(f(a)) = f^{-1}(a) = i(a) = a$. Thus, since $a = f^{-1}(f(a)) = f^{-1}(a)$, we have that $f^{-1} \in H(a)$. Moreover, H is nonempty. (Why?) Consequently, $H(a)$ is a subgroup of G.

4. Let G be as in Example 6 of Section 1, and H as in Example 7. Then H is a subgroup of G (see Problem 3 in Section 1).

5. Let G be as in Example 6, H as in Example 7, and K as in Example 8 in Section 1. Then $K \subset H \subset G$ and K is a subgroup of both H and of G.

6. Let $\mathbb{C}'$ be the nonzero complex numbers as a group under the multiplication of complex numbers. Let $V = \{a \in \mathbb{C}' \mid |a| \text{ is rational}\}$. Then V is a subgroup of $\mathbb{C}'$. For if $|a|$ and $|b|$ are rational, then $|ab| = |a|\,|b|$ is rational, so

$ab \in V$; also, $|a^{-1}| = 1/|a|$ is rational, hence $a^{-1} \in V$. Therefore, V is a subgroup of $\mathbb{C}'$.

7. Let $\mathbb{C}'$ and V be as above and let

$$U = \{a \in \mathbb{C}' \mid a = \cos\theta + i\sin\theta, \theta \text{ any real}\}.$$

If $a = \cos\theta + i\sin\theta$ and $b = \cos\psi + i\sin\psi$, we saw in Chapter 1 that $ab = \cos(\theta + \psi) + i\sin(\theta + \psi)$, so that $ab \in U$, and that $a^{-1} = \cos\theta - i\sin\theta = \cos(-\theta) + i\sin(-\theta) \in U$. Also, $|a| = 1$, since $|a| = \sqrt{\cos^2\theta + \sin^2\theta} = 1$. Therefore, $U \subset V \subset \mathbb{C}'$ and U is a subgroup both of V and of $\mathbb{C}'$.

8. Let $\mathbb{C}'$, U, V be as above, and let $n > 1$ be an integer. Let $\theta_n = \cos(2\pi/n) + i\sin(2\pi/n)$, and let $B = \{1, \theta_n, \theta_n^2, \ldots, \theta_n^{n-1}\}$. Since $\theta_n^n = 1$ (as we saw by De Moivre's Theorem), it is easily checked that B is a subgroup of U, V, and $\mathbb{C}'$, and is of order n.

9. Let G be any group and let $a \in G$. The set $A = \{a^i \mid i \text{ any integer}\}$ is a subgroup of G. For, by the rules of exponents, if $a^i \in A$ and $a^j \in A$, then $a^i a^j = a^{i+j}$, so is in A. Also, $(a^i)^{-1} = a^{-i}$, so $(a^i)^{-1} \in A$. This makes A into a subgroup of G.

A is the cyclic subgroup of G generated by a in the following sense.

Definition. The *cyclic subgroup of* G generated by a is a set $\{a^i \mid i \text{ any integer}\}$. It is denoted (a).

Note that if e is the identity element of G, then $(e) = \{e\}$. In Example 8, the group B is the cyclic group (θ_n) of $\mathbb{C}$ generated by θ_n.

10. Let G be any group; for $a \in G$ let $C(a) = \{g \in G \mid ga = ag\}$. We claim that $C(a)$ is a subgroup of G. First, the closure of $C(a)$. If $g, h \in C(a)$, then $ga = ag$ and $ha = ah$, thus $(gh)a = g(ha) = g(ah) = (ga)h = (ag)h = a(gh)$ (by the repeated use of the associative law), hence $gh \in C(a)$. Also, if $g \in C(a)$, then from $ga = ag$ we have $g^{-1}(ga)g^{-1} = g^{-1}(ag)g^{-1}$, which simplifies to $ag^{-1} = g^{-1}a$; whence $g^{-1} \in C(a)$. So, $C(a)$ is thereby a subgroup of G.

These particular subgroups $C(a)$ will come up later for us and they are given a special name. We call $C(a)$ the *centralizer* of a in G. If in a group $ab = ba$, we say that a and b *commute*. Thus $C(a)$ is the set of all elements in G that commute with a.

11. Let G be any group and let $Z(G) = \{z \in G \mid zx = xz \text{ for all } x \in G\}$. We leave it to the reader to verify that $Z(G)$ is a subgroup of G. It is called the *center* of G.

12. Let G be any group and H a subgroup of G. For $a \in G$, let $a^{-1}Ha = \{a^{-1}ha \mid h \in H\}$. We assert that $a^{-1}Ha$ is a subgroup of G. If $x = a^{-1}h_1a$ and $y = a^{-1}h_2a$ where $h_1, h_2 \in H$, then $xy = (a^{-1}h_1a)(a^{-1}h_2a) = a^{-1}(h_1h_2)a$ (associative law), and since H is a subgroup of G, $h_1h_2 \in H$. Therefore, $a^{-1}(h_1h_2)a \in a^{-1}Ha$, which says that $xy \in a^{-1}Ha$. Thus $a^{-1}Ha$ is closed. Also, if $x = a^{-1}ha \in a^{-1}Ha$, then, as is easily verified, $x^{-1} = (a^{-1}ha)^{-1} = a^{-1}h^{-1}a \in a^{-1}Ha$. Therefore, $a^{-1}Ha$ is a subgroup of G.

An even dozen seems to be about the right number of examples, so we go on to other things. Lemma 2.3.1 points out for us what we need in order that a given subset of a group be a subgroup. In an important special case we can make a considerable saving in checking whether a given subset H is a subgroup of G. This is the case in which H is *finite*.

Lemma 2.3.2. Suppose that G is a group and H a nonempty *finite* subset of G closed under the product in G. Then H is a subgroup of G.

Proof. By Lemma 2.3.1 we must show that $a \in H$ implies $a^{-1} \in H$. If $a = e$, then $a^{-1} = e$ and we are done. Suppose then that $a \neq e$; consider the elements $a, a^2, \ldots, a^{n+1}$, where $n = |H|$, the order of H. *Here we have written down $n + 1$ elements, all of them in H since H is closed, and H has only n distinct elements.* How can this be? Only if some two of the elements listed are equal; put another way, only if $a^i = a^j$ for some $1 \leq i < j \leq n + 1$. But then, by the cancellation property in groups, $a^{j-i} = e$. Since $j - i \geq 1$, $a^{j-i} \in H$, hence $e \in H$. However, $j - i - 1 \geq 0$, so $a^{j-i-1} \in H$ and $aa^{j-i-1} = a^{j-i} = e$, whence $a^{-1} = a^{j-i-1} \in H$. This proves the lemma. □

An immediate, but nevertheless important, corollary to Lemma 2.3.2 is the

Corollary. If G is a finite group and H a nonempty subset of G closed under multiplication, then H is a subgroup of G.

PROBLEMS

Easier Problems

1. If A, B are subgroups of G, show that $A \cap B$ is a subgroup of G.
2. What is the cyclic subgroup of $\mathbb{Z}$ generated by -1 under $+$?
3. Let S_3 be the symmetric group of degree 3. Find all the subgroups of S_3.
4. Verify that $Z(G)$, the center of G, is a subgroup of G. (See Example 11.)

5. If $C(a)$ is the centralizer of a in G (Example 10), prove that $Z(G) = \cap_{a \in G} C(a)$.
6. Show that $a \in Z(G)$ if and only if $C(a) = G$.
7. In S_3, find $C(a)$ for each $a \in S_3$.
8. If G is an abelian group and if $H = \{a \in G \mid a^2 = e\}$, show that H is a subgroup of G.
9. Give an example of a nonabelian group for which the H in Problem 8 is *not* a subgroup.
10. If G is an abelian group and $n > 1$ an integer, let $A_n = \{a^n \mid a \in G\}$. Prove that A_n is a subgroup of G.
*11. If G is an abelian group and $H = \{a \in G \mid a^{n(a)} = e$ for some $n(a) > 1$ depending on $a\}$, prove that H is a subgroup of G.

We say that a group G is *cyclic* if there exists an $a \in G$ such that every $x \in G$ is a power of a, that is, $x = a^j$ for some j. In other words, G is cyclic if $G = (a)$ for some $a \in G$, in which case we say that a is a *generator* for G.

*12. Prove that a cyclic group is abelian.
13. If G is cyclic, show that every subgroup of G is cyclic.
14. If G has no proper subgroups, prove that G is cyclic.
15. If G is a group and H a nonempty subset of G such that, given $a, b \in H$, then $ab^{-1} \in H$, prove that H is a subgroup of G.

Middle-Level Problems

*16. If G has no proper subgroups, prove that G is cyclic of order p, where p is a prime number. (This sharpens the result of Problem 14.)
17. If G is a group and $a, x \in G$, prove that $C(x^{-1}ax) = x^{-1}C(a)x$. [See Examples 10 and 12 for the definitions of $C(b)$ and of $x^{-1}C(a)x$.]
18. If S is a nonempty set and $X \subset S$, show that $T(X) = \{f \in A(S) \mid f(X) \subset X\}$ is a subgroup of $A(S)$ *if X is finite.*
19. If A, B are subgroups of an abelian group G, let $AB = \{ab \mid a \in A, b \in B\}$. Prove that AB is a subgroup of G.
20. Give an example of a group G and two subgroups A, B of G such that AB is *not* a subgroup of G.
21. If A, B are subgroups of G such that $b^{-1}Ab \subset A$ for all $b \in B$, show that AB is a subgroup of G.
*22. If A and B are finite subgroups, of orders m and n, respectively, of the abelian group G, prove that AB is a subgroup of order mn *if m and n are relatively prime.*

23. What is the order of AB in Problem 22 if m and n are not relatively prime?

24. If H is a subgroup of G, let $N = \cap_{x \in G} x^{-1}Hx$. Prove that N is a subgroup of G such that $y^{-1}Ny = N$ for every $y \in G$.

Harder Problems

25. Let S, X, $T(X)$ be as in Problem 18 (but X no longer finite). Give an example of a set S and an infinite subset X such that $T(X)$ is *not* a subgroup of $A(S)$.

***26.** Let G be a group, H a subgroup of G. Let $Hx = \{hx \mid h \in H\}$. Show that, given $a, b \in G$, then $Ha = Hb$ or $Ha \cap Hb = \emptyset$.

***27.** If in Problem 26 H is a finite subgroup of G, prove that Ha and Hb have the same number of elements. What is this number?

28. Let M, N be subgroups of G such that $x^{-1}Mx \subset M$ and $x^{-1}Nx \subset N$ for all $x \in G$. Prove that MN is a subgroup of G and that $x^{-1}(MN)x \subset MN$ for all $x \in G$.

***29.** If M is a subgroup of G such that $x^{-1}Mx \subset M$ for all $x \in G$, prove that actually $x^{-1}Mx = M$.

30. If M, N are such that $x^{-1}Mx = M$ and $x^{-1}Nx = N$ for all $x \in G$, and if $M \cap N = (e)$, prove that $mn = nm$ for any $m \in M$, $n \in N$. (**Hint:** Consider the element $m^{-1}n^{-1}mn$.)

4. LAGRANGE'S THEOREM

We are about to derive the first real group-theoretic result of importance. Although its proof is relatively easy, this theorem is like the A-B-C's for finite groups and has interesting implications in number theory.

As a matter of fact, those of you who solved Problems 26 and 27 of Section 3 have all the necessary ingredients to effect a proof of the result. The theorem simply states that in a finite group the order of a subgroup divides the order of the group.

To smooth the argument of this theorem—which is due to Lagrange—and for use many times later, we make a short detour into the realm of set theory.

Just as the concept of "function" runs throughout most phases of mathematics, so also does the concept of "relation." A *relation* is a statement aRb about the elements $a, b \in S$. If S is the set of integers, $a = b$ is a relation on S. Similarly, $a < b$ is a relation on S, as is $a \leq b$.

Definition. A relation $\sim$ on a set S is called an *equivalence relation* if, for all $a, b, c \in S$, it satisfies:

(a) $a \sim a$ (*reflexivity*).
(b) $a \sim b$ implies that $b \sim a$ (*symmetry*).
(c) $a \sim b$, $b \sim c$ implies that $a \sim c$ (*transitivity*).

Of course, equality, $=$, is an equivalence relation, so the general notion of equivalence relation is a generalization of that of equality. In a sense, an equivalence relation measures equality with regard to some attribute. This vague remark may become clearer after we see some examples.

Examples

1. Let S be all the items for sale in a grocery store; we declare $a \sim b$, for $a, b \in S$, if the price of a equals that of b. Clearly, the defining rules of an equivalence relation hold for this $\sim$. Note that in measuring this "generalized equality" on S we ignore all properties of the elements of S other than their price. So $a \sim b$ if they are equal as far as the attribute of price is concerned.

2. Let S be the integers and $n > 1$ a fixed integer. We define $a \sim b$ for $a, b \in S$ if $n \mid (a - b)$. We verify that this is an equivalence relation. Since $n \mid 0$ and $0 = a - a$, we have $a \sim a$. Because $n \mid (a - b)$ implies that $n \mid (b - a)$, we have that $a \sim b$ implies that $b \sim a$. Finally, if $a \sim b$ and $b \sim c$, then $n \mid (a - b)$ and $n \mid (b - c)$; hence $n \mid ((a - b) + (b - c))$, that is, $n \mid (a - c)$. Therefore, $a \sim c$.

This relation on the integers is of great importance in number theory and is called *congruence modulo* n; when $a \sim b$, we write this as $a \equiv b \bmod n$ [or, sometimes, as $a \equiv b(n)$], which is read "a congruent to b mod n." We'll be running into it very often from now on. As we shall see, this is a special case of a much wider phenomenon in groups.

3. We generalize Example 2. Let G be a group and H a subgroup of G. For $a, b \in G$, define $a \sim b$ if $ab^{-1} \in H$. Since $e \in H$ and $e = aa^{-1}$, we have that $a \sim a$. Also, if $ab^{-1} \in H$, then since H is a subgroup of G, $(ab^{-1})^{-1} \in H$. But $(ab^{-1})^{-1} = (b^{-1})^{-1}a^{-1} = ba^{-1}$, so $ba^{-1} \in H$, hence $b \sim a$. This tells us that $a \sim b$ implies that $b \sim a$. Finally, if $a \sim b$ and $b \sim c$, then $ab^{-1} \in H$ and $bc^{-1} \in H$. But $(ab^{-1})(bc^{-1}) = ac^{-1}$, whence $ac^{-1} \in H$ and therefore $a \sim c$. We have shown the transitivity of $\sim$, thus $\sim$ is an equivalence relation on G.

Note that if $G = \mathbb{Z}$, the group of integers under $+$, and H is the subgroup consisting of all multiples of n, for $n > 1$ a fixed integer, then $ab^{-1} \in H$

translates into $a \equiv b(n)$. So congruence mod n is a very special case of the equivalence we have defined in Example 3.

It is this equivalence relation that we shall use in proving Lagrange's theorem.

4. Let G be any group. For $a, b \in G$ we declare that $a \sim b$ if there exists an $x \in G$ such that $b = x^{-1}ax$. We claim that this defines an equivalence relation on G. First, $a \sim a$ for $a = e^{-1}ae$. Second, if $a \sim b$, then $b = x^{-1}ax$, hence $a = (x^{-1})^{-1}b(x^{-1})$, so that $b \sim a$. Finally, if $a \sim b$, $b \sim c$, then $b = x^{-1}ax$, $c = y^{-1}by$ for some $x, y \in G$. Thus $c = y^{-1}(x^{-1}ax)y = (xy)^{-1}a(xy)$, and so $a \sim c$. We have established that this defines an equivalence relation on G.

This relation, too, plays an important role in group theory and is given the special name *conjugacy*. When $a \sim b$ we say that "*a and b are conjugate in G.*" Note that if G is abelian, then $a \sim b$ if and only if $a = b$.

We could go on and on to give numerous interesting examples of equivalence relations, but this would sidetrack us from our main goal in this section. There will be no lack of examples in the problems at the end of this section.

We go on with our discussion and make the

Definition. If $\sim$ is an equivalence relation on S, then $[a]$, the *class* of a, is defined by $[a] = \{b \in S \mid b \sim a\}$.

Let us see what the class of a is in the two examples, Examples 3 and 4, just given.

In Example 3, $a \sim b$ if $ab^{-1} \in H$, that is, if $ab^{-1} = h$, for some $h \in H$. Thus $a \sim b$ implies that $a = hb$. On the other hand, if $a = kb$ where $k \in H$, then $ab^{-1} = (kb)b^{-1} = k \in H$, so $a \sim b$ if and only if $a \in Hb = \{hb \mid h \in H\}$. Therefore, $[b] = Hb$.

The set Hb is called a *right coset of H* in G. We ran into such in Problem 26 of Section 3. Note that $b \in Hb$, since $b = eb$ and $e \in H$ (also because $b \in [b] = Hb$). Right cosets, and left handed counterparts of them called *left cosets*, play important roles in what follows.

In Example 4, we defined $a \sim b$ if $b = x^{-1}ax$ for some $x \in G$. Thus $[a] = \{x^{-1}ax \mid x \in G\}$. We shall denote $[a]$ in this case as $\mathrm{cl}(a)$ and call it the *conjugacy class* of a in G. If G is abelian, then $\mathrm{cl}(a)$ consists of a alone. In fact, if $a \in Z(G)$, the center of G, then $\mathrm{cl}(a)$ consists merely of a.

The notion of conjugacy and its properties will crop up again often, especially in Section 11.

We shall examine the class of an element a in Example 2 later in this chapter.

The important influence that an equivalence relation has on a set is to break it up and *partition* it into nice disjoint pieces.

Theorem 2.4.1. If $\sim$ is an equivalence relation on S, then $S = \cup[a]$, where this union runs over one element from each class, and where $[a] \neq [b]$ implies that $[a] \cap [b] = \emptyset$. That is, $\sim$ partitions S into equivalence classes.

Proof. Since $a \in [a]$, we have $\cup_{a \in S}[a] = S$. The proof of the second assertion is also quite easy. We show that if $[a] \neq [b]$, then $[a] \cap [b] = \emptyset$, or, what is equivalent to this, if $[a] \cap [b] \neq \emptyset$, then $[a] = [b]$.

Suppose, then, that $[a] \cap [b] \neq \emptyset$; let $c \in [a] \cap [b]$. By definition of class, $c \sim a$ since $c \in [a]$ and $c \sim b$ since $c \in [b]$. Therefore, $a \sim c$ by symmetry of $\sim$, and so, since $a \sim c$ and $c \sim b$, we have $a \sim b$. Thus $a \in [b]$; if $x \in [a]$, then $x \sim a$, $a \sim b$ gives us that $x \sim b$, hence $x \in [b]$. Thus $[a] \subset [b]$. The argument is obviously symmetric in a and b, so we have $[b] \subset [a]$, whence $[a] = [b]$, and our assertion above is proved.

The theorem is now completely proved. □

We now can prove a famous result of Lagrange.

Theorem 2.4.2 (Lagrange's Theorem). If G is a finite group and H is a subgroup of G, then the order of H divides the order of G.

Proof. Let us look back at Example 3, where we established that the relation $a \sim b$ if $ab^{-1} \in H$ is an equivalence relation and that

$$[a] = Ha = \{ha \mid h \in H\}.$$

Let k be the number of distinct classes—call them $Ha_1, \ldots, Ha_k$. By Theorem 2.4.1, $G = Ha_1 \cup Ha_2 \cup \cdots \cup Ha_k$ and we know that $Ha_j \cap Ha_i = \emptyset$ if $i \neq j$.

We assert that any Ha_i has $|H|$ = order of H number of elements. Map $H \to Ha_i$ by sending $h \to ha_i$. We claim that this map is 1-1, for if $ha_i = h'a_i$, then by cancellation in G we would have $h = h'$; thus the map is 1-1. It is definitely onto by the very definition of Ha_i. So H and Ha_i have the same number, $|H|$, of elements.

Since $G = Ha_1 \cup \cdots \cup Ha_k$ and the Ha_i are disjoint and each Ha_i has $|H|$ elements, we have that $|G| = k|H|$. Thus $|H|$ divides $|G|$ and Lagrange's Theorem is proved. □

Although Lagrange sounds like a French name, J. L. Lagrange (1736–1813) was actually Italian, having been born and brought up in Turin. He spent most of his life, however, in France. Lagrange was a great mathematician who made fundamental contributions to all the areas of mathematics of his day.

If G is finite, the number of right cosets of H in G, namely $|G|/|H|$, is called the *index* of H in G and is written as $i_G(H)$.

Recall that a group G is said to be *cyclic* if there is an element $a \in G$ such that every element in G is a power of a.

Theorem 2.4.3. A group G of prime order is cyclic.

Proof. If H is a subgroup of G then, by invoking Lagrange's Theorem, $|H|$ divides $|G| = p$, p a prime, so $|H| = 1$ or p. So if $H \neq (e)$, then $H = G$. If $a \in G$, $a \neq e$, then the powers of a form a subgroup (a) of G different from (e). So this subgroup is all of G. This says that any $x \in G$ is of the form $x = a^i$. Hence, G is cyclic by the definition of cyclic group. □

If G is finite and $a \in G$, we saw earlier in the proof of Lemma 2.3.2 that $a^{n(a)} = e$ for some $n(a) \geq 1$, depending on a. We make the

Definition. If G is finite, then the *order* of a, written $o(a)$, is the *least positive integer* m such that $a^m = e$.

Suppose that $a \in G$ has order m. Consider the set $A = \{e, a, a^2, \ldots, a^{m-1}\}$; we claim that A is a subgroup of G (since $a^m = e$) and that the m elements listed in A are distinct. We leave the verification of these claims to the reader. Thus $|A| = m = o(a)$. Since $|A| \mid |G|$, we have

Theorem 2.4.4. If G is finite and $a \in G$, then $o(a) \mid |G|$.

If $a \in G$, where G is finite, we have, by Theorem 2.4.4, $|G| = k \cdot o(a)$. Thus

$$a^{|G|} = a^{k \cdot o(a)} = (a^{o(a)})^k = e^k = e.$$

We have proved the

Theorem 2.4.5. If G is a finite group of order n, then $a^n = e$ for all $a \in G$.

When we apply this last result to certain special groups arising in number theory, we shall obtain some classical number-theoretic results due to Fermat and Euler.

Let $\mathbb{Z}$ be the integers and let $n > 1$ be a fixed integer. We go back to Example 2 of equivalence relations, where we defined $a \equiv b \bmod n$ (a congruent to $b \bmod n$) if $n \mid (a - b)$. The class of a, $[a]$, consists of all $a + nk$, where k runs through all the integers. We call it the *congruence class* of a.

By Euclid's Algorithm, given any integer b, $b = qn + r$, where $0 \leq r < n$, thus $[b] = [r]$. So the n classes $[0], [1], \ldots, [n-1]$ give us all the congruence classes. We leave it to the reader to verify that they are distinct.

Let $\mathbb{Z}_n = \{[0], [1], \ldots, [n-1]\}$. We shall introduce two operations, $+$ and $\cdot$ in $\mathbb{Z}_n$. Under $+$ $\mathbb{Z}_n$ will form an abelian group; under $\cdot$ $\mathbb{Z}_n$ will not form a group, but a certain piece of it will become a group.

How to define $[a] + [b]$? What is more natural than to define

$$[a] + [b] = [a + b].$$

But there is a fly in the ointment. Is this operation $+$ in $\mathbb{Z}_n$ well-defined? What does that mean? We can represent $[a]$ by many a's—for instance, if $n = 3$, $[1] = [4] = [-2] = \cdots$, yet we are using a *particular* a to define the addition. What we must show is that if $[a] = [a']$ and $[b] = [b']$, then $[a + b] = [a' + b']$, for then we will have $[a] + [b] = [a + b] = [a' + b'] = [a'] + [b']$.

Suppose that $[a] = [a']$; then $n \mid (a - a')$. Also from $[b] = [b']$, $n \mid (b - b')$, hence $n \mid ((a - a') + (b - b')) = ((a + b) - (a' + b'))$. Therefore, $a + b \equiv (a' + b') \bmod n$, and so $[a + b] = [a' + b']$.

So we now have a well-defined addition in $\mathbb{Z}_n$. The element $[0]$ acts as the identity element and $[-a]$ acts as $-[a]$, the inverse of $[a]$. We leave it to the reader to check out that $\mathbb{Z}_n$ is a group under $+$. It is a cyclic group of order n generated by $[1]$.

We summarize this all as

Theorem 2.4.6. $\mathbb{Z}_n$ forms a cyclic group under the addition $[a] + [b] = [a + b]$.

Having disposed of the addition in $\mathbb{Z}_n$, we turn to the introduction of a multiplication. Again, what is more natural than defining

$$[a] \cdot [b] = [ab]?$$

So, for instance, if $n = 9$, $[2][7] = [14] = [5]$, and $[3][6] = [18] = [0]$. Under this multiplication—we leave the fact that it is well-defined to the reader—$\mathbb{Z}_n$ *does not form* a group. Since $[0][a] = [0]$ for all a, and the unit element under multiplication is $[1]$, $[0]$ cannot have a multiplicative inverse. Okay, why not try the nonzero elements $[a] \neq [0]$ as a candidate for a group under this product? Here again it is no go *if n is not a prime.* For instance, if $n = 6$, then $[2] \neq [0]$, $[3] \neq [0]$, yet $[2][3] = [6] = [0]$, so the nonzero elements do not, in general, give us a group.

So we ask: Can we find an appropriate piece of $\mathbb{Z}_n$ that will form a

group under multiplication? Yes! Let $U_n = \{[a] \in \mathbb{Z}_n \mid (a, n) = 1\}$, noting that $(a, n) = 1$ if and only if $(b, n) = 1$ for $[a] = [b]$. By the Corollary to Theorem 1.5.5, if $(a, n) = 1$ and $(b, n) = 1$, then $(ab, n) = 1$. So $[a][b] = [ab]$ yields that if $[a], [b] \in U_n$, then $[ab] \in U_n$ and U_n is closed. Associativity is easily checked, following from the associativity of the integers under multiplication. The identity element is easy to find, namely [1]. Multiplication is commutative in U_n.

Note that if $[a][b] = [a][c]$ where $[a] \in U_n$, then we have $[ab] = [ac]$, and so $[ab - ac] = [0]$. This says that $n \mid a(b - c) = ab - ac$; but a is relatively prime to n. By Theorem 1.5.5 one must have that $n \mid (b - c)$, and so $[b] = [c]$. In other words, we have the cancellation property in U_n. By Problem 2 of Section 2, U_n is a group.

What is the order of U_n? By the definition of U_n, $|U_n|$ = number of integers $1 \leq m < n$ such that $(m, n) = 1$. This number comes up often and we give it a name.

Definition. The *Euler φ-function*, $\varphi(n)$, is defined by $\varphi(1) = 1$ and, for $n > 1$, $\varphi(n)$ = the number of positive integers m with $1 \leq m < n$ such that $(m, n) = 1$.

Thus $|U_n| = \varphi(n)$. If $n = p$, a prime, we have $\varphi(p) = p - 1$. We see that $\varphi(8) = 4$ for only 1, 3, 5, 7 are less than 8 and positive and relatively prime to 8. We try another one, $\varphi(15)$. The numbers $1 \leq m < 15$ relatively prime to 15 are 1, 2, 4, 7, 8, 11, 13, 14, so $\varphi(15) = 8$.

Let us look at some examples of U_n.

1. $U_8 = \{[1], [3], [5], [7]\}$. Note that $[3][5] = [15] = [7]$, $[5]^2 = [25] = [1]$. In fact, U_8 is a group of order 4 in which $a^2 = e$ for every $a \in U_8$.

2. $U_{15} = \{[1], [2], [4], [7], [8], [11], [13], [14]\}$. Note that $[11][13] = [143] = [8]$, $[2]^4 = [1]$, and so on.

The reader should verify that $a^4 = e = [1]$ for every $a \in U_{15}$.

3. $U_9 = \{[1], [2], [4], [5], [7], [8]\}$. Note that $[2]^1 = [2]$, $[2]^2 = [4]$, $[2]^3 = [8]$, $[2^4] = [16] = [7]$, $[2]^5 = [32] = [5]$; also $[2]^6 = [2][2]^5 = [2][5] = [10] = [1]$. So the powers of [2] give us every element in U_9. Thus U_9 is a *cyclic* group of order 6. What other elements in U_9 generate U_9?

In parallel to Theorem 2.4.6 we have

Theorem 2.4.7. U_n forms an abelian group, under the product $[a][b] = [ab]$, of order $\varphi(n)$, where $\varphi(n)$ is the Euler φ-function.

An immediate consequence of Theorems 2.4.7 and 2.4.5 is a famous result in number theory.

Theorem 2.4.8 (Euler). If a is an integer relatively prime to n, then $a^{\varphi(n)} \equiv 1 \bmod n$.

Proof. U_n forms a group of order $\varphi(n)$, so by Theorem 2.4.5, $a^{\varphi(n)} = e$ for all $a \in U_n$. This translates into $[a^{\varphi(n)}] = [a]^{\varphi(n)} = [1]$, which in turn translates into $n \mid (a^{\varphi(n)} - 1)$ for every integer a relatively prime to p. In other words, $a^{\varphi(n)} \equiv 1 \bmod n$. □

A special case, where $n = p$ is a prime, is due to Fermat.

Corollary (Fermat). If p is a prime and $p \nmid a$, then

$$a^{p-1} \equiv 1 \bmod p.$$

For any integer b, $b^p \equiv b \bmod p$.

Proof. Since $\varphi(p) = p - 1$, if $(a, p) = 1$, we have, by Theorem 2.4.8, that $a^{p-1} \equiv 1(p)$, hence $a^1 \cdot a^{p-1} \equiv a(p)$, so that $a^p \equiv a(p)$. If $p \mid b$, then $b \equiv 0(p)$ and $b^p \equiv 0(p)$, so that $b^p \equiv b(p)$. □

Leonard Euler (1707–1785) was probably the greatest scientist that Switzerland has produced. He was the most prolific of all mathematicians ever.

Pierre Fermat (1601–1665) was a great number theorist. Fermat's Last Theorem—which was in fact first proved in 1994 by Andrew Wiles—states that the equation $a^n + b^n = c^n$ (a, b, c, n being integers) has only the trivial solution where $a = 0$ or $b = 0$ or $c = 0$ if $n > 2$.

One final cautionary word about Lagrange's Theorem. Its *converse* in general is *not* true. That is, if G is a finite group of order n, then it need not be true that for every divisor m of n there is a subgroup of G of order m. A group with this property is very special indeed, and its structure can be spelled out quite well and precisely.

PROBLEMS

Easier Problems

1. Verify that the relation $\sim$ is an equivalence relation on the set S given.

(a) $S = \mathbb{R}$ reals, $a \sim b$ if $a - b$ is rational.

(b) $S = \mathbb{C}$, the complex numbers, $a \sim b$ if $|a| = |b|$.

(c) $S =$ straight lines in the plane, $a \sim b$ if a, b are parallel.

(d) $S =$ set of all people, $a \sim b$ if they have the same color eyes.

2. The relation $\sim$ on the real numbers $\mathbb{R}$ defined by $a \sim b$ if both $a > b$ and $b > a$ is *not* an equivalence relation. Why not? What properties of an equivalence relation does it satisfy?

3. Let $\sim$ be a relation on a set S that satisfies (1) $a \sim b$ implies that $b \sim a$ and (2) $a \sim b$ and $b \sim c$ implies that $a \sim c$. These seem to imply that $a \sim a$. For if $a \sim b$, then by (1), $b \sim a$, so $a \sim b$, $b \sim a$, so by (2), $a \sim a$. If this argument is correct, then the relation $\sim$ must be an equivalence relation. Problem 2 shows that this is not so. What is wrong with the argument we have given?

4. Let S be a set, $\{S_\alpha\}$ nonempty subsets such that $S = \cup_\alpha S_\alpha$ and $S_\alpha \cap S_\beta = \emptyset$ if $\alpha \neq \beta$. Define an equivalence relation on S in such a way that the S_α are precisely all the equivalence classes.

*** 5.** Let G be a group and H a subgroup of G. Define, for $a, b \in G$, $a \sim b$ if $a^{-1}b \in H$. Prove that this defines an equivalence relation on G, and show that $[a] = aH = \{ah \mid h \in H\}$. The sets aH are called *left cosets* of H in G.

6. If G is S_3 and $H = \{i, f\}$, where $f: S \to S$ is defined by $f(x_1) = x_2, f(x_2) = x_1, f(x_3) = x_3$, list all the right cosets of H in G and list all the left cosets of H in G.

7. In Problem 6, is every right coset of H in G also a left coset of H in G?

8. If every right coset of H in G is a left coset of H in G, prove that $aHa^{-1} = H$ for all $a \in G$.

9. In $\mathbb{Z}_{16}$, write down all the cosets of the subgroup $H = \{[0], [4], [8], [12]\}$. (Since the operation in $\mathbb{Z}_n$ is $+$, write your coset as $[a] + H$. We don't need to distinguish between right cosets and left cosets, since $\mathbb{Z}_n$ is abelian under $+$.)

10. In Problem 9, what is the index of H in $\mathbb{Z}_{16}$? (Recall that we defined the index $i_G(H)$ as the number of right cosets in G.)

11. For any finite group G, show that there are as many distinct left cosets of H in G as there are right cosets of H in G.

12. If aH and bH are distinct left cosets of H in G, are Ha and Hb distinct right cosets of H in G? Prove that this is true or give a counterexample.

13. Find the orders of all the elements of U_{18}. Is U_{18} cyclic?

14. Find the orders of all the elements of U_{20}. Is U_{20} cyclic?

***15.** If p is a prime, show that the only solutions of $x^2 \equiv 1 \bmod p$ are $x \equiv 1 \bmod p$ or $x \equiv -1 \bmod p$.

***16.** If G is a finite abelian group and $a_1, \ldots, a_n$ are all its elements, show that $x = a_1 a_2 \cdots a_n$ must satisfy $x^2 = e$.

17. If G is of odd order, what can you say about the x in Problem 16?

18. Using the results of Problems 15 and 16, prove that if p is an odd prime number, then $(p-1)! \equiv -1 \bmod p$. (This is known as *Wilson's Theorem.*) It is, of course, also true if $p = 2$.

19. Find all the distinct conjugacy classes of S_3.

20. In the group G of Example 6 of Section 1, find the conjugacy class of the element $T_{a,b}$. Describe it in terms of a and b.

21. Let G be the dihedral group of order 8 (see Example 9, Section 1). Find the conjugacy classes in G.

22. Verify Euler's Theorem for $n = 14$ and $a = 3$, and for $n = 14$ and $a = 5$.

23. In U_{41}, show that there is an element a such that $[a]^2 = [-1]$, that is, an integer a such that $a^2 \equiv -1 \bmod 41$.

24. If p is a prime number of the form $4n + 3$, show that we *cannot* solve

$$x^2 \equiv -1 \bmod p$$

[**Hint:** Use Fermat's Theorem that $a^{p-1} \equiv 1 \bmod p$ if $p \nmid a$.]

25. Show that the nonzero elements in $\mathbb{Z}_n$ form a group under the product $[a][b] = [ab]$ if and only if n is a prime.

Middle-Level Problems

26. Let G be a group, H a subgroup of G, and let S be the set of all distinct right cosets of H in G, T the set of all left cosets of H in G. Prove that there is a 1-1 mapping of S onto T. (**Note:** The obvious map that comes to mind, which sends Ha into aH, is not the right one. See Problems 5 and 12.)

27. If $aH = bH$ forces $Ha = Hb$ in G, show that $aHa^{-1} = H$ for every $a \in G$.

28. If G is a cyclic group of order n, show that there are $\varphi(n)$ generators for G. Give their form explicitly.

29. If in a group G, $aba^{-1} = b^i$, show that $a^r b a^{-r} = b^{i^r}$ for all positive integers r.

30. If in G $a^5 = e$ and $aba^{-1} = b^2$, find $o(b)$ if $b \neq e$.

***31.** If $o(a) = m$ and $a^s = e$, prove that $m \mid s$.

32. Let G be a finite group, H a subgroup of G. Let $f(a)$ be the least positive m such that $a^m \in H$. Prove that $f(a) \mid o(a)$.

33. If $i \neq f \in A(S)$ is such that $f^p = i$, p a prime, and if for some $s \in S$, $f^j(s) = s$ for some $1 \leq j < p$, show that $f(s) = s$.

34. If $f \in A(S)$ has order p, p a prime, show that for every $s \in S$ the orbit of s under f has one or p elements. [**Recall:** The orbit of s under f is $\{f^j(s) \mid j \text{ any integer}\}$.]

35. If $f \in A(S)$ has order p, p a prime, and S is a finite set having n elements, where $(n, p) = 1$, show that for some $s \in S$, $f(s) = s$.

Harder Problems

36. If $a > 1$ is an integer, show that $n \mid \varphi(a^n - 1)$, where φ is the Euler φ-function. [**Hint:** Consider the integers mod$(a^n - 1)$.]

37. In a cyclic group of order n, show that for each positive integer m that divides n (including $m = 1$ and $m = n$) there are $\varphi(m)$ elements of order m.

38. Using the result of Problem 37, show that $n = \Sigma_{m \mid n} \varphi(m)$.

39. Let G be a finite abelian group of order n for which the number of solutions of $x^m = e$ is at most m for any m dividing n. Prove that G must be cyclic. [**Hint:** Let $\psi(m)$ be the number of elements in G of *order* m. Show that $\psi(m) \leq \varphi(m)$ and use Problem 38.]

40. Using the result of Problem 39, show that U_p, if p is a prime, is cyclic. (This is a famous result in number theory; it asserts the existence of a *primitive root* mod p.)

41. Using the result of Problem 40, show that if p is a prime of the form $p = 4n + 1$, then we can solve $x^2 \equiv -1 \bmod p$ (with x an integer).

42. Using Wilson's Theorem (see Problem 28), show that if p is a prime of the form $p = 4n + 1$ and if

$$y = 1 \cdot 2 \cdot 3 \cdots \frac{p-1}{2} = \left(\frac{p-1}{2}\right)!,$$

then $y^2 \equiv -1 \bmod p$. (This gives another proof of the result in Problem 41.)

43. Let G be an abelian group of order n, and $a_1, \ldots, a_n$ its elements. Let $x = a_1 a_2 \cdots a_n$. Show that:

(a) If G has exactly one element $b \neq e$ such that $b^2 = e$, then $x = b$.

(b) If G has more than one element $b \neq e$ such that $b^2 = e$, then $x = e$.

(c) If n is odd, then $x = e$ (see Problem 16).

5. HOMOMORPHISMS AND NORMAL SUBGROUPS

In a certain sense the subject of group theory is built up out of three basic concepts: that of a homomorphism, that of a normal subgroup, and that of the factor or quotient group of a group by a normal subgroup. We discuss the first two of these in this section, and the third in Section 6.

Without further ado we introduce the first of these.

Definition. Let G, G' be two groups; then the mapping $\varphi: G \to G'$ is a *homomorphism* if $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in G$.

(*Note:* This φ has nothing to do with the Euler φ-function.)

In this definition the product on the left side—in $\varphi(ab)$—is that of G, while the product $\varphi(a)\varphi(b)$ is that of G'. A short description of a homomorphism is that it *preserves* the operation of G. We do *not* insist that φ be onto; if it is, we'll say that it is. Before working out some facts about homomorphisms, we present some examples.

Examples

1. Let G be the group of all positive reals under the multiplication of reals, and G' the group of all reals under addition. Let $\varphi: G \to G'$ be defined by $\varphi(x) = \log_{10} x$ for $x \in G$. Since $\log_{10}(xy) = \log_{10} x + \log_{10} y$, we have $\varphi(xy) = \varphi(x) + \varphi(y)$, so φ is a homomorphism. It also happens to be onto and 1-1.

2. Let G be an *abelian* group and let $\varphi: G \to G$ be defined by $\varphi(a) = a^2$. Since $\varphi(ab) = (ab)^2 = a^2b^2 = \varphi(a)\varphi(b)$, φ is a homomorphism of G into itself. It need not be onto; the reader should check that in U_8 (see Section 4) $a^2 = e$ for all $a \in U_8$, so $\varphi(G) = (e)$.

3. The example of U_8 above suggests the so-called *trivial homomorphism.* Let G be any group and G' any other; define $\varphi(x) = e'$, the unit element of G', for all $x \in G$. Trivially, φ is a homomorphism of G into G'. It certainly is not a very interesting one.

Another homomorphism always present is the identity mapping, i, of any group G into itself. Since $i(x) = x$ for all $x \in G$, clearly $i(xy) = xy = i(x)i(y)$. The map i is 1-1 and onto, but, again, is not too interesting as a homomorphism.

4. Let G be the group of integers under $+$ and $G' = \{1, -1\}$, the subgroup of the reals under multiplication. Define $\varphi(m) = 1$ if m is even, $\varphi(m) = -1$ if m is odd. The statement that φ is a homomorphism is merely a restatement of:

even + even = even, even + odd = odd, and odd + odd = even.

5. Let G be the group of all nonzero complex numbers under multiplication and let G' be the group of positive reals under multiplication. Let $\varphi: G \to G'$ be defined by $\varphi(a) = |a|$; then $\varphi(ab) = |ab| = |a|\,|b| = \varphi(a)\varphi(b)$, so φ is a homomorphism of G into G'. In fact, φ is onto.

6. Let G be the group in Example 6 of Section 1, and G' the group of nonzero reals under multiplication. Define $\varphi: G \to G'$ by $\varphi(T_{a,b}) = a$. That

φ is a homomorphism follows from the product rule in G, namely, $T_{a,b}T_{c,d} = T_{ac,ad+b}$.

7. Let $G = \mathbb{Z}$ be the group of integers under $+$ and let $G' = \mathbb{Z}_n$. Define $\varphi: G \to \mathbb{Z}_n$ by $\varphi(m) = [m]$. Since the addition in $\mathbb{Z}_n$ is defined by $[m] + [r] = [m + r]$, we see that $\varphi(m + r) = \varphi(m) + \varphi(r)$, so φ is indeed a homomorphism of $\mathbb{Z}$ onto $\mathbb{Z}_n$.

8. The following general construction gives rise to a well-known theorem. Let G be any group, and let $A(G)$ be the set of all 1-1 mappings of G onto itself—here we are viewing G merely as a set, forgetting about its multiplication. Define $T_a: G \to G$ by $T_a(x) = ax$ for every $x \in G$. *What is the product, T_aT_b, of T_a and T_b as mappings on G?* Well,

$$(T_aT_b)(x) = T_a(T_bx) = T_a(bx) = a(bx) = (ab)x = T_{ab}(x)$$

(we used the associative law). So we see that $T_aT_b = T_{ab}$.

Define the mapping $\varphi: G \to A(G)$ by $\varphi(a) = T_a$, for $a \in G$. The product rule for the T's translates into $\varphi(ab) = T_{ab} = T_aT_b = \varphi(a)\varphi(b)$, so φ is a homomorphism of G into $A(G)$. We claim that φ is 1-1. Suppose that $\varphi(a) = \varphi(b)$, that is, $T_a = T_b$. Therefore, $a = T_a(e) = T_b(e) = b$, so φ is indeed 1-1. It is not onto in general—for instance, if G has order $n > 2$, then $A(G)$ has order $n!$, and since $n! > n$, φ doesn't have a ghost of a chance of being onto. It is easy to verify that the image of φ, $\varphi(G) = \{T_a \mid a \in G\}$, is a subgroup of $A(G)$.

The fact that φ is 1-1 suggests that perhaps 1-1 homomorphisms should play a special role. We single them out in the following definition.

Definition. The homomorphism $\varphi: G \to G'$ is called a *monomorphism* if φ is 1-1. A monomorphism that is onto is called an *isomorphism*. An isomorphism from G to G itself is called an *automorphism*.

One more definition.

Definition. Two groups G and G' are said to be *isomorphic* if there is an isomorphism of G onto G'.

We shall denote that G and G' are isomorphic by writing $G \simeq G'$.

This definition seems to be asymmetric, but, in point of fact, it is not. For if there is an isomorphism of G onto G', there is one of G' onto G (see Problem 2).

We shall discuss more thoroughly later what it means for two groups to be isomorphic. But now we summarize what we did in Example 8.

Theorem 2.5.1 (Cayley's Theorem). Every group G is isomorphic to some subgroup of $A(S)$, for an appropriate S.

The appropriate S we used was G itself. But there may be better choices. We shall see some in the problems to follow.

When G is finite, we can take the set S in Theorem 2.5.1 to be finite, in which case $A(S)$ is S_n and its elements are permutations. In this case, Cayley's Theorem is usually stated as: *A finite group can be represented as a group of permutations.*

(Arthur Cayley (1821–1895) was an English mathematician who worked in matrix theory, invariant theory, and many other parts of algebra.)

This is a good place to discuss the importance of "isomorphism." Let φ be an isomorphism of G onto G'. We can view G' as a relabeling of G, using the label $\varphi(x)$ for the element x. Is this labeling consistent with the structure of G as a group? That is, if x is labeled $\varphi(x)$, y labeled $\varphi(y)$, what is xy labeled as? Since $\varphi(x)\varphi(y) = \varphi(xy)$, we see that xy is labeled as $\varphi(x)\varphi(y)$, so this renaming of the elements is consistent with the product in G. So two groups that are isomorphic—although they need not be equal—in a certain sense, as described above, are equal. Often, it is desirable to be able to identify a given group as isomorphic to some concrete group that we know.

We go on with more examples.

9. Let G be any group, $a \in G$ fixed in the discussion. Define $\varphi: G \to G$ by $\varphi(x) = a^{-1}xa$ for all $x \in G$. We claim that φ is an isomorphism of G onto itself. First,

$$\varphi(xy) = a^{-1}(xy)a = a^{-1}xa \cdot a^{-1}ya = \varphi(x)\varphi(y),$$

so φ is at least a homomorphism of G into itself. It is 1-1 for if $\varphi(x) = \varphi(y)$, then $a^{-1}xa = a^{-1}ya$, so by cancellation in G we get $x = y$. Finally, φ is onto, for $x = a^{-1}(axa^{-1})a = \varphi(axa^{-1})$ for any $x \in G$.

Here φ is called the *inner automorphism* of G *induced* by a. The notion of *automorphism* and some of its properties will come up in the problems.

One final example:

10. Let G be the group of reals under $+$ and let G' be the group of all nonzero complex numbers under multiplication. Define $\varphi: G \to G'$ by

$$\varphi(x) = \cos x + i \sin x.$$

We saw that $(\cos x + i \sin x)(\cos y + i \sin y) = \cos(x + y) + i \sin(x + y)$,

hence $\varphi(x)\varphi(y) = \varphi(x + y)$ and φ is a homomorphism of G into G'. φ is not 1-1 because, for instance, $\varphi(0) = \varphi(2\pi) = 1$, nor is φ onto.

Now that we have a few examples in hand, we start a little investigation of homomorphisms. We begin with

Lemma 2.5.2. If φ is a homomorphism of G into G', then:

(a) $\varphi(e) = e'$, the unit element of G'.

(b) $\varphi(a^{-1}) = \varphi(a)^{-1}$ for all $a \in G$.

Proof. Since $x = xe$, $\varphi(x) = \varphi(xe) = \varphi(x)\varphi(e)$; by cancellation in G' we get $\varphi(e) = e'$. Also, $\varphi(aa^{-1}) = \varphi(e) = e'$, hence $e' = \varphi(aa^{-1}) = \varphi(a)\varphi(a^{-1})$, which proves that $\varphi(a^{-1}) = \varphi(a)^{-1}$. □

Definition. The *image* of φ, $\varphi(G)$, is $\varphi(G) = \{\varphi(a) \mid a \in G\}$.

We leave to the reader the proof of

Lemma 2.5.3. If φ is a homomorphism of G into G', then the image of φ is a subgroup of G'.

We singled out certain homomorphisms and called them monomorphisms. Their property was that they were 1-1. We want to measure how far a given homomorphism is from being a monomorphism. This prompts the

Definition. If φ is a homomorphism of G into G', then the *kernel* of φ, Ker φ, is defined by Ker $\varphi = \{a \in G \mid \varphi(a) = e'\}$.

Ker φ measures the lack of *1-1' ness* at one point e'. We claim that this lack is rather uniform. What is $W = \{x \in G \mid \varphi(x) = w'\}$ for a given $w' \in G'$? We show that if $\varphi(x) = w'$ for some $x \in G$, then $W = \{kx \mid k \in \text{Ker } \varphi\} = (\text{Ker } \varphi)x$. Clearly, if $k \in \text{Ker } \varphi$ and $\varphi(x) = w'$, then $\varphi(kx) = \varphi(k)\varphi(x) = e'\varphi(x) = w'$, so $kx \in W$. Also, if $\varphi(x) = \varphi(y) = w'$, then $\varphi(x) = \varphi(y)$, hence $\varphi(y)\varphi(x)^{-1} = e'$; but $\varphi(x)^{-1} = \varphi(x^{-1})$ by Lemma 2.5.2, so $e' = \varphi(y)\varphi(x)^{-1} = \varphi(y)\varphi(x^{-1}) = \varphi(yx^{-1})$, whence $yx^{-1} \in \text{Ker } \varphi$ and so $y \in (\text{Ker } \varphi)x$. *Thus the inverse image of any element* w' in $\varphi(G) \in G'$ is the set $(\text{Ker } \varphi)x$, where x is *any* element in G such that $\varphi(x) = w'$.

We state this as

Lemma 2.5.4. If $w' \in G'$ is of the form $\varphi(x) = w'$, then $\{y \in G \mid \varphi(y) = w'\} = (\text{Ker } \varphi)x$.

We now shall study some basic properties of the kernels of homomorphisms.

Theorem 2.5.5. If φ is a homomorphism of G into G', then

(a) Ker φ is a subgroup of G.

(b) Given $a \in G$, a^{-1} (Ker φ)$a \subset$ Ker φ.

Proof. Although this is so important, its proof is easy. If $a, b \in$ Ker φ, then $\varphi(a) = \varphi(b) = e'$, hence $\varphi(ab) = \varphi(a)\varphi(b) = e'$, whence $ab \in$ Ker φ, so Ker φ is closed under product. Also $\varphi(a) = e'$ implies that $\varphi(a^{-1}) = \varphi(a)^{-1} = e'$, and so $a^{-1} \in$ Ker φ. Therefore, Ker φ is a subgroup of G. If $k \in$ Ker φ and $a \in G$, then $\varphi(k) = e'$. Consequently, $\varphi(a^{-1}ka) = \varphi(a^{-1})\varphi(k)\varphi(a) = \varphi(a^{-1})e'\varphi(a) = \varphi(a^{-1})\varphi(a) = \varphi(a^{-1}a) = \varphi(e) = e'$. This tells us that $a^{-1}ka \in$ Ker φ, hence a^{-1} (Ker φ) $a \in$ Ker φ. The theorem is now completely proved. □

Corollary. If φ is a homomorphism of G into G', then φ is a monomorphism if and only if Ker $\varphi = (e)$.

Proof. This result is really a corollary to Lemma 2.5.4. We leave the few details to the reader. □

Property (b) of Ker φ in Theorem 2.5.5 is an interesting and basic one for a subgroup to enjoy. We ran into this property in the text material and problems earlier on several occasions. We use it to define the ultra-important class of subgroups of a group.

Definition. The subgroup N of G is said to be a *normal subgroup* of G if $a^{-1}Na \subset N$ for every $a \in G$.

Of course, Ker φ, for any homomorphism, is a normal subgroup of G. As we shall see in the next section, every normal subgroup of G is the kernel of some appropriate homomorphism of G into an appropriate group G'. So in a certain sense the notions of homomorphism and normal subgroups will be shown to be equivalent.

Although we defined a normal subgroup via $a^{-1}Na \subset N$, we actually have $a^{-1}Na = N$. For if $a^{-1}Na \subset N$ for all $a \in G$, then $N = a(a^{-1}Na)a^{-1} \subset aNa^{-1} = (a^{-1})^{-1}Na^{-1} \subset N$. So $N = aNa^{-1}$ for every $a \in G$. Transposing, we have $Na = aN$; that is, every left coset of N in G is a right coset of N in G.

On the other hand, if every left coset of N in G is a right coset, then the left coset aN, which contains a, must be equal to the right coset containing a,

namely Na. Thus, $aN = Na$ and $N = a^{-1}Na$ for all $a \in G$, which is to say that N is normal in G.

We write "N is a normal subgroup of G" by the abbreviated symbol $N \lhd G$.

Note that $a^{-1}Na = N$ *does not* mean that $a^{-1}na = n$ for every $n \in N$. No—merely that the set of all $a^{-1}na$ is the same as the set of all n.

We have proved

Theorem 2.5.6. $N \lhd G$ if and only if every left coset of N in G is a right coset of N in G.

Before going any further, we pause to look at some examples of kernels of homomorphisms and normal subgroups.

If G is abelian, then every subgroup of G is normal, for $a^{-1}xa = x$ for every $a, x \in G$. The converse of this is *not* true. Nonabelian groups exist in which *every subgroup* is normal. See if you can find such an example of order 8. Such nonabelian groups are called *Hamiltonian*, after the Irish mathematician W. R. Hamilton (1805–1865). The desired group of order 8 can be found in the *quaternions* of Hamilton, which we introduce in Chapter 4, Section 1.

In Example 1, $\varphi(x) = \log_{10} x$, and Ker $\varphi = \{x \mid \log_{10} x = 0\} = \{1\}$. In Example 2, where G is abelian, and $\varphi(x) = x^2$,

$$\text{Ker } \varphi = \{x \in G \mid x^2 = e\}$$

The kernel of the trivial homomorphism of Example 3 is *all* of G. In Example 4, Ker φ is the set of all even integers. In Example 5, Ker $\varphi = \{a \in C' \mid |a| = 1\}$, which can be identified, from the polar form of a complex number, as Ker $\varphi = \{\cos x + i \sin x \mid x \text{ real}\}$. In Example 6, Ker $\varphi = \{T_{1,b} \in G \mid b \text{ real}\}$. In Example 7, Ker φ is the set of all multiples of n. In Examples 8 and 9, the kernels consists of e alone, for the maps are monomorphisms. In Example 10, we see that Ker $\varphi = \{2\pi m \mid m \text{ any integer}\}$.

Of course, all the kernels above are normal subgroups of their respective groups. We should look at some normal subgroups, intrinsically in G itself, without recourse to the kernels of homomorphism. We go back to the examples of Section 1.

1. In Example 7, $H = \{T_{a,b} \in G \mid a \text{ rational}\}$. If $T_{x,y} \in G$, we leave it to the reader to check that $T_{x,y}^{-1} H T_{x,y} \subset H$ and so $H \lhd G$.
2. In Example 9 the subgroup $\{i, g, g^2, g^3\} \lhd G$. Here too we leave the checking to the reader.
3. In Example 10 the subgroup $H = \{i, h, h^2, \ldots, h^{n-1}\}$ is normal in G. This we also leave to the reader.

4. If G is any group, $Z(G)$, the *center* of G, is a normal subgroup of G (see Example 11 of Section 3).
5. If $G = S_3$, G has the elements i, f, g, g^2, fg, and gf, where $f(x_1) = x_2$, $f(x_2) = x_1$, $f(x_3) = x_3$ and $g(x_1) = x_2$, $g(x_2) = x_3$, $g(x_3) = x_1$. We claim that the subgroup $N = \{i, g, g^2\} \lhd S_3$. As we saw earlier (or can compute now), $fgf^{-1} = g^{-1} = g^2$, $fg^2f^{-1} = g$. $(fg)\, g(fg)^{-1} = fggg^{-1}f^{-1} = fgf^{-1} = g^2$, and so on. So $N \lhd S_3$ follows.

The material in this section has been a rather rich diet. It may not seem so, but the ideas presented, although simple, are quite subtle. We recommend that the reader digest the concepts and results thoroughly before going on. One way of seeing how complete this digestion is, is to take a stab at many of the almost infinite list of problems that follow. The material of the next section is even a richer diet, and even harder to digest. Avoid a mathematical stomachache later by assimilating this section well.

PROBLEMS

Easier Problems

1. Determine in each of the parts if the given mapping is a homomorphism. If so, identify its kernel and whether or not the mapping is 1-1 or onto.
- **(a)** $G = \mathbb{Z}$ under $+$, $G' = \mathbb{Z}_n$, $\varphi(a) = [a]$ for $a \in \mathbb{Z}$.
- **(b)** G group, $\varphi: G \to G$ defined by $\varphi(a) = a^{-1}$ for $a \in G$.
- **(c)** G abelian group, $\varphi: G \to G$ defined by $\varphi(a) = a^{-1}$ for $a \in G$.
- **(d)** G group of all nonzero real numbers under multiplication, $G' = \{1, -1\}$, $\varphi(r) = 1$ if r is positive, $\varphi(r) = -1$ if r is negative.
- **(e)** G an abelian group, $n > 1$ a fixed integer, and $\varphi: G \to G$ defined by $\varphi(a) = a^n$ for $a \in G$.

2. Recall that $G \simeq G'$ means that G is isomorphic to G'. Prove that for all groups G_1, G_2, G_3:
- **(a)** $G_1 \simeq G_1$.
- **(b)** $G_1 \simeq G_2$ implies that $G_2 \simeq G_1$.
- **(c)** $G_1 \simeq G_2$, $G_2 \simeq G_3$ implies that $G_1 \simeq G_3$.

3. Let G be any group and $A(G)$ the set of all 1-1 mappings of G, as a set, onto itself. Define $L_a: G \to G$ by $L_a(x) = xa^{-1}$. Prove that:
- **(a)** $L_a \in A(G)$.
- **(b)** $L_aL_b = L_{ab}$.
- **(c)** The mapping $\psi: G \to A(G)$ defined by $\psi(a) = L_a$ is a monomorphism of G into $A(G)$.

4. In Problem 3 prove that for all $a, b \in G$, $T_a L_b = L_b T_a$, where T_a is defined as in Example 8.
5. In Problem 4, show that if $V \in A(G)$ is such that $T_a V = V T_a$ for all $a \in G$, then $V = L_b$ for some $b \in G$. (**Hint:** Acting on $e \in G$, find out what b should be.)
6. Prove that if $\varphi: G \to G'$ is a homomorphism, then $\varphi(G)$, the image of G, is a subgroup of G'.
7. Show that $\varphi: G \to G'$, where φ is a homomorphism, is a monomorphism if and only if Ker $\varphi = (e)$.
8. Find an isomorphism of G, the group of all real numbers under $+$, *onto* G', the group of all positive real numbers under multiplication.
9. Verify that if G is the group in Example 6 of Section 1, and $H = \{T_{a,b} \in G \mid a \text{ rational}\}$, then $H \lhd G$, the dihedral group of order 8.
10. Verify that in Example 9 of Section 1, the set $H = \{i, g, g^2, g^3\}$ is a normal subgroup of G, the dihedral group of order 8.
11. Verify that in Example 10 of Section 1, the subgroup

$$H = \{i, h, h^2, \ldots, h^{n-1}\}$$

is normal in G.

12. Prove that if $Z(G)$ is the center of G, then $Z(G) \lhd G$.
13. If G is a finite abelian group of order n and $\varphi: G \to G$ is defined by $\varphi(a) = a^m$ for all $a \in G$, find the necessary and sufficient condition that φ be an isomorphism of G onto itself.
14. If G is abelian and $\varphi: G \to G'$ is a homomorphism of G *onto* G', prove that G' is abelian.
15. If G is any group, $N \lhd G$, and $\varphi: G \to G'$ a homomorphism of G *onto* G', prove that the image, $\varphi(N)$, of N is a normal subgroup of G'.
16. If $N \lhd G$ and $M \lhd G$ and $MN = \{mn \mid m \in M, n \in N\}$, prove that MN is a subgroup of G and that $MN \lhd G$.
17. If $M \lhd G$, $N \lhd G$, prove that $M \cap N \lhd G$.
18. If H is any subgroup of G and $N = \cap_{a \in G} a^{-1} H a$, prove that $N \lhd G$.
19. If H is a subgroup of G, let $N(H)$ be defined by the relation $N(H) = \{a \in G \mid a^{-1} H a = H\}$. Prove that:
    (a) $N(H)$ is a subgroup of G and $N(H) \supset H$.
    (b) $H \lhd N(H)$.
    (c) If K is a subgroup of G such that $H \lhd K$, then $K \subset N(H)$. [So $N(H)$ is the largest subgroup of G in which H is normal.]
20. If $M \lhd G$, $N \lhd G$, and $M \cap N = (e)$, show that for $m \in M, n \in N$, $mn = nm$.

21. Let S be any set having more than two elements and $A(S)$ the set of all 1-1 mappings of S onto itself. If $s \in S$, we define $H(s) = \{f \in A(S) \mid f(s) = s\}$. Prove that $H(s)$ *cannot* be a normal subgroup of $A(S)$.

22. Let $G = S_3$, the symmetric group of degree 3 and let $H = \{i, f\}$, where $f(x_1) = x_2, f(x_2) = x_1, f(x_3) = x_3$.

(a) Write down all the left cosets of H in G.

(b) Write down all the right cosets of H in G.

(c) Is every left coset of H a right coset of H?

23. Let G be a group such that all subgroups of G are normal in G. If $a, b \in G$, prove that $ba = a^j b$ for some j.

24. If G_1, G_2 are two groups, let $G = G_1 \times G_2$, the Cartesian product of G_1, G_2 [i.e., G is the set of all ordered pairs (a, b) where $a \in G_1$, $b \in G_2$]. Define a product in G by $(a_1, b_1)(a_2, b_2) = (a_1a_2, b_1b_2)$.

(a) Prove that G is a group.

(b) Show that there is a monomorphism φ_1 of G_1 into G such that $\varphi_1(G_1) \lhd G$, given by $\varphi_1(a_1) = (a_1, e_2)$, where e_2 is the identity element of G_2.

(c) Find the similar monomorphism φ_2 of G_2 into G.

(d) Using the mappings φ_1, φ_2 of Parts (b) and (c), prove that $\varphi_1(G_1)\varphi_2(G_2) = G$ and $\varphi_1(G_1) \cap \varphi_2(G_2)$ is the identity element of G.

(e) Prove that $G_1 \times G_2 \simeq G_2 \times G_1$.

25. Let G be a group and let $W = G \times G$ as defined in Problem 24. Prove that:

(a) The mapping $\varphi: G \to W$ defined by $\varphi(a) = (a, a)$ is a monomorphism of G into W.

(b) The image $\varphi(G)$ in W [i.e., $\{(a, a) \mid a \in G\}$] is normal in W if and only if G is abelian.

Middle-Level Problems

***26.** If G is a group and $a \in G$, define $\sigma_a: G \to G$ by $\sigma_a(g) = aga^{-1}$. We saw in Example 9 of this section that σ_a is an isomorphism of G onto itself, so $\sigma_a \in A(G)$, the group of all 1-1 mappings of G (as a set) onto itself. Define $\psi: G \to A(G)$ by $\psi(a) = \sigma_a$ for all $a \in G$. Prove that:

(a) ψ is a homomorphism of G into $A(G)$.

(b) Ker $\psi = Z(G)$, the center of G.

27. If θ is an automorphism of G and $N \lhd G$, prove that $\theta(N) \lhd G$.

28. Let θ, ψ be automorphisms of G, and let $\theta\psi$ be the product of θ and ψ as mappings on G. Prove that $\theta\psi$ is an automorphism of G, and that θ^{-1} is an automorphism of G, so that the set of all automorphisms of G is itself a group.

*29. A subgroup T of a group W is called *characteristic* if $\varphi(T) \subset T$ for all automorphisms, φ, of W. Prove that:
   (a) M characteristic in G implies that $M \triangleleft G$.
   (b) M, N characteristic in G implies that MN is characteristic in G.
   (c) A normal subgroup of a group *need not* be characteristic. (This is quite hard; you must find an example of a group G and a noncharacteristic normal subgroup.)

30. Suppose that $|G| = pm$, where $p \nmid m$ and p is a prime. If H is a normal subgroup of order p in G, prove that H is characteristic.

31. Suppose that G is an abelian group of order $p^n m$ where $p \nmid m$ is a prime. If H is a subgroup of G of order p^n, prove that H is a characteristic subgroup of G.

32. Do Problem 31 even if G is not abelian if you happen to know that for some reason or other $H \triangleleft G$.

33. Suppose that $N \triangleleft G$ and $M \subset N$ is a characteristic subgroup of N. Prove that $M \triangleleft G$. (It is *not* true that if $M \triangleleft N$ and $N \triangleleft G$, then M must be normal in G. See Problem 50.)

34. Let G be a group, $\mathcal{A}(G)$ the group of all automorphisms of G. (See Problem 28.) Let $I(G) = \{\sigma_a \mid a \in G\}$, where σ_a is as defined in Problem 26. Prove that $I(G) \triangleleft \mathcal{A}(G)$.

35. Show that $Z(G)$, the center of G, is a characteristic subgroup of G.

36. If $N \triangleleft G$ and H is a subgroup of G, show that $H \cap N \triangleleft H$.

Harder Problems

37. If G is a nonabelian group of order 6, prove that $G \simeq S_3$.

38. Let G be a group and H a subgroup of G. Let $S = \{Ha \mid a \in G\}$ be the set of all right cosets of H in G. Define, for $b \in G$, $T_b : S \to S$ by $T_b(Ha) = Hab^{-1}$.
   (a) Prove that $T_b T_c = T_{bc}$ for all $b, c \in G$ [therefore the mapping $\psi : G \to A(S)$ defined by $\psi(b) = T_b$ is a homomorphism].
   (b) Describe Ker ψ, the kernel of $\psi : G \to A(S)$.
   (c) Show that Ker ψ is the largest normal subgroup of G lying in H [largest in the sense that if $N \triangleleft G$ and $N \subset H$, then $N \subset$ Ker ψ].

39. Use the result of Problem 38 to redo Problem 37.

Recall that if H is a subgroup of G, then the *index* of H in G, $i_G(H)$, is the number of distinct right cosets of H and G (if this number is finite).

40. If G is a finite group, H a subgroup of G such that $n \nmid i_G(H)!$ where $n = |G|$, prove that there is a normal subgroup $N \neq (e)$ of G contained in H.

41. Suppose that you know that a group G of order 21 contains an element a of order 7. Prove that $A = (a)$, the subgroup generated by a, is normal in G. (**Hint:** Use the result of Problem 40.)

42. Suppose that you know that a group G of order 36 has a subgroup H or order 9. Prove that either $H \lhd G$ or there exists a subgroup $N \lhd G$, $N \subset H$, and $|N| = 3$.

43. Prove that a group of order 9 must be abelian.

44. Prove that a group of order p^2, p a prime, has a normal subgroup of order p.

45. Using the result of Problem 44, prove that a group of order p^2, p a prime, must be abelian.

46. Let G be a group of order 15; show that there is an element $a \neq e$ in G such that $a^3 = e$ and an element $b \neq e$ such that $b^5 = e$.

47. In Problem 46, show that both subgroups $A = \{e, a, a^2\}$ and $B = \{e, b, b^2, b^3, b^4\}$ are normal in G.

48. From the result of Problem 47, show that any group of order 15 is cyclic.

Very Hard Problems

49. Let G be a group, H a subgroup of G such that $i_G(H)$ is finite. Prove that there is a subgroup $N \subset H$, $N \lhd G$ such that $i_G(N)$ is finite.

50. Construct a group G such that G has a normal subgroup N, and N has a normal subgroup M (i.e., $N \lhd G$, $M \lhd N$), yet M is not normal in G.

51. Let G be a finite group, φ an automorphism of G such that φ^2 is the identity automorphism of G. Suppose that $\varphi(x) = x$ implies that $x = e$. Prove that G is abelian and $\varphi(a) = a^{-1}$ for all $a \in G$.

52. Let G be a finite group and φ an automorphism of G such that $\varphi(x) = x^{-1}$ for *more than three-fourths* of the elements of G. Prove that $\varphi(y) = y^{-1}$ for *all* $y \in G$, and so G is abelian.

6. FACTOR GROUPS

Let G be a group and N a normal subgroup of G. In proving Lagrange's Theorem we used, for an arbitrary subgroup H, the equivalence relation $a \sim b$ if $ab^{-1} \in H$. Let's try this out when N is normal and see if we can say a little more than one could say for just any old subgroup.

So, let $a \sim b$ if $ab^{-1} \in N$ and let $[a] = \{x \in G \mid x \sim a\}$. As we saw earlier, $[a] = Na$, the right coset of N in G containing a. Recall that in looking at $\mathbb{Z}_n$ we defined for it an operation $+$ via $[a] + [b] = [a + b]$. Why

not try something similar for an arbitrary group G and a normal subgroup N of G?

So let $M = \{[a] \mid a \in G\}$, where $[a] = \{x \in G \mid xa^{-1} \in N\} = Na$. We define a product in M via $[a][b] = [ab]$. We shall soon show that M is a group under this product. *But first and foremost we must show that this product in M is well-defined.* In other words, we must show that if $[a] = [a']$ and $[b] = [b']$, then $[ab] = [a'b']$, for this would show that $[a][b] = [ab] = [a'b'] = [a'][b']$; equivalently, that *this product of classes does not depend on the particular representatives we use for the classes.*

Therefore let us suppose that $[a] = [a']$ and $[b] = [b']$. From the definition of our equivalence we have that $a' = na$, where $n \in N$. Similarly, $b' = mb$, where $m \in N$. Thus $a'b' = namb = n(ama^{-1})ab$; since $N \lhd G$, ama^{-1} is in N, so $n(ama^{-1})$ is also in N. So if we let $n_1 = n(ama^{-1})$, then $n_1 \in N$ and $a'b' = n_1ab$. But this tells us that $a'b' \in Nab$, so that $a'b' \sim ab$, from which we have that $[a'b'] = [ab]$, the exact thing we required to ensure that our product in M was well-defined.

Thus M is now endowed with a well-defined product $[a][b] = [ab]$. We now verify the group axioms for M. Closure we have from the very definition of this product. If $[a]$, $[b]$, and $[c]$ are in M, then $[a]([b][c]) = [a][bc] = [a(bc)] = [(ab)c]$ (since the product in G is associative) $= [ab][c] = ([a][b])[c]$. Therefore, the associative law has been established for the product in M. What about a unit element? Why not try the obvious choice, namely $[e]$? We immediately see that $[a][e] = [ae] = [a]$ and $[e][a] = [ea] = [a]$, so $[e]$ does act as the unit element for M. Finally, what about inverses? Here, too, the obvious choice is the correct one. If $a \in G$, then $[a][a^{-1}] = [aa^{-1}] = [e]$, hence $[a^{-1}]$ acts as the inverse of $[a]$ relative to the product we have defined in M.

We want to give M a name, and better still, a symbol that indicates its dependence on G and N. The symbol we use for M is G/N (read "G over N or G mod N") and G/N is called the *factor group* or *quotient group* of G by N.

What we have shown is the very important

Theorem 2.6.1. If $N \lhd G$ and

$$G/N = \{[a] \mid a \in G\} = \{Na \mid a \in G\},$$

then G/N is a group relative to the operation $[a][b] = [ab]$.

One observation must immediately be made, namely

Theorem 2.6.2. If $N \lhd G$, then there is a homomorphism ψ of G onto G/N such that Ker ψ, the kernel of ψ, is N.

Proof. The most natural mapping from G to G/N is the one that does the trick. Define $\psi: G \to G/N$ by $\psi(a) = [a]$. Our product as defined in G/N makes of ψ a homomorphism, for $\psi(ab) = [ab] = [a][b] = \psi(a)\psi(b)$. Since every element $X \in G/N$ is of the form $X = [b] = \psi(b)$ for some $b \in G$, ψ is onto. Finally, what is the kernel, Ker ψ, of ψ? By definition, Ker $\psi = \{a \in G \mid \psi(a) = E\}$, where E is the unit element of G/N. But what is E? Nothing other than $E = [e] = Ne = N$, and $a \in$ Ker ψ if and only if $E = N = \psi(a) = Na$. But $Na = N$ tells us that $a = ea \in Na = N$, so we see that Ker $\psi \subset N$. That $N \subset$ Ker ψ—which is easy—we leave to the reader. So Ker $\psi = N$. □

Theorem 2.6.2 substantiates the remark we made in the preceding section that every normal subgroup N of G is the kernel of some homomorphism of G onto some group. The "some homomorphism" is the ψ defined above and the "some group" is G/N.

This construction of the factor group G by N is possibly the single most important construction in group theory. In other algebraic systems we shall have analogous constructions, as we shall see later.

One might ask: Where in this whole affair did the normality of N in G enter? Why not do the same thing for any subgroup H of G? So let's try and see what happens. As before, we define

$$W = \{[a] \mid a \in G\} = \{Ha \mid a \in G\}$$

where the equivalence $a \sim b$ is defined by $ab^{-1} \in H$. We try to introduce a product in W as we did for G/N by defining $[a][b] = [ab]$. Is this product well defined? If $h \in H$, then $[hb] = [b]$, so for the product to be well defined, we would need that $[a][b] = [a][hb]$, that is, $[ab] = [ahb]$. This gives us that $Hab = Hahb$, and so $Ha = Hah$; this implies that $H = Haha^{-1}$, whence $aha^{-1} \in H$. That is, for all $a \in G$ and all $h \in H$, aha^{-1} must be in H; in other words, H must be normal in G. *So we see that in order for the product defined in W to be well-defined, H must be a normal subgroup of G.*

We view this matter of the quotient group in a slightly different way. If A, B are subsets of G, let $AB = \{ab \mid a \in A, b \in B\}$. If H is a subgroup of G, then $HH \subset H$ is another way of saying that H is closed under the product of G.

Let $G/N = \{Na \mid a \in G\}$ be the set of all right cosets of the normal subgroup N in G. Using the product of subsets of G as defined above, what is $(Na)(Nb)$? By definition, $(Na)(Nb)$ consists of all elements of the form $(na)(mb)$, where $n, m \in N$, and so

$$(na)(mb) = (nama^{-1})(ab) = n_1ab,$$

where $n_1 = nama^{-1}$ is in N, since N is normal. Thus $(Na)(Nb) \subset Nab$. On the other hand, if $n \in N$, then

$$n(ab) = (na)(eb) \in (Na)(Nb),$$

so that $Nab \subset (Na)(Nb)$. In short, we have shown that the product—as subsets of G—of Na and Nb is given by the formula $(Na)(Nb) = Nab$. All the other group axioms for G/N, as defined here, are now readily verified from this product formula.

Another way of seeing that $(Na)(Nb) = Nab$ is to note that by the normality of N, $aN = Na$, hence $(Na)(Nb) = N(aN)b = N(Na)b = NNab = Nab$, since $NN = N$ (because N is a subgroup of G).

However we view G/N—as equivalence classes or as a set of certain subsets of G—we do get a group whose structure is intimately tied to that of G, via the natural homomorphism ψ of G onto G/N.

We shall see very soon how we combine induction and the structure of G/N to get information about G.

When G is a finite group and $N \lhd G$, then the number of right cosets of N in G, $i_G(N)$, is given—as the proof of Lagrange's Theorem showed—by $i_G(n) = |G|/|N|$. But this is the order of G/N, which is the set of all the right cosets of N in G. Thus $|G/N| = |G|/|N|$. We state this more formally as

Theorem 2.6.3. If G is a finite group and $N \lhd G$, then $|G/N| = |G|/|N|$.

As an application of what we have been talking about here, we shall prove a special case of a theorem that we shall prove in its full generality later. The proof we give—for the abelian case—is not a particularly good one, but it illustrates quite clearly a general technique, that of pulling back information about G/N to get information about G itself.

The theorem we are about to prove is due to the great French mathematician A. L. Cauchy (1789–1857), whose most basic contributions were in complex variable theory.

Theorem 2.6.4 (Cauchy). If G is a finite abelian group of order $|G|$ and p is a prime that divides $|G|$, then G has an element of order p.

Proof. Before getting involved with the proof, we point out to the reader that the theorem is true for *any* finite group. We shall prove it in the general case later, with a proof that will be much more beautiful than the one we are about to give for the special, abelian case.

We proceed by induction on $|G|$. What does this mean precisely? We shall

assume the theorem to be true for all abelian groups of order less than $|G|$ and show that this forces the theorem to be true for G. If $|G| = 1$, there is no such p and the theorem is vacuously true. So we have a starting point for our induction.

Suppose that there is a subgroup $(e) \neq N \neq G$. Since $|N| < |G|$, if p divides $|N|$, by our induction hypothesis there would be an element of order p in N, hence in G, and we would be done. So we may suppose that $p \nmid |N|$. Since G is abelian, every subgroup is normal, so we can form G/N. Because p divides $|G|$ and $p \nmid |N|$, and because $|G/N| = |G|/|N|$, we have that p divides $|G/N|$. The group G/N is abelian, since G is (Prove!) and since $N \neq (e)$, $|N| > 1$, so $|G/N| = |G|/|N| < |G|$. Thus, again by induction, there exists an element in G/N of order p. In other words, there exists an $a \in G$ such that $[a]^p = [e]$, but $[a] \neq [e]$. This translates to $a^p \in N$, $a \notin N$. So if $m = |N|$, then $(a^p)^m = e$. So $(a^m)^p = e$. If we could show that $b = a^m \neq e$, then b would be the required element of order p in G. But if $a^m = e$, then $[a]^m = [e]$, and since $[a]$ has order p, $p \mid m$ (see Problem 31 of Section 4). But, by assumption, $p \nmid m = |N|$. *So we are done if G has a nontrivial subgroup.*

But if G has no nontrivial subgroups, it must be cyclic of prime order. (See Problem 16 of Section 3, which you should be able to handle more easily now.) What is this "prime order"? Because p divides $|G|$, we must have $|G| = p$. But then any element $a \neq e \in G$ satisfies $a^p = e$ and is of order p. This completes the induction, and so proves the theorem. □

We shall have other applications of this kind of group-theoretic argument in the problems.

The notion of a factor group is a very subtle one, and of the greatest importance in the subject. The formation of a new set from an old one by using as elements of this new set subsets of the old one is strange to the neophyte seeing this kind of construction for the first time. So it is worthwhile looking at this whole matter from a variety of points of view. We consider G/N from another angle now.

What are we doing when we form G/N? Sure, we are looking at equivalences classes defined via N. Let's look at it another way. What we are doing is *identifying* two elements in G if they satisfy the relation $ab^{-1} \in N$. In a sense we are blotting out N. So although G/N is *not* a subgroup of G, we can look at it as G, with N blotted out, and two elements as equal if they are equal "up to N."

For instance, in forming $\mathbb{Z}/N$, where $\mathbb{Z}$ is the group of integers and N is the set of all multiples of 5 in $\mathbb{Z}$, what we are doing is *identifying* 1 with 6, 11, 16, -4, -9, and so on, and we are identifying all multiples of 5 with 0. The nice thing about all this is that this identification jibes with addition in $\mathbb{Z}$ when we go over to $\mathbb{Z}/N$.

Let's look at a few examples from this point of view.

1. Let $G = \{T_{a,b} \mid a \neq 0, b \text{ real}\}$ (Example 6 of Section 1). Let $N = \{T_{1,b} \mid b \text{ real}\} \subset G$; we saw that $N \lhd G$, so it makes sense to talk about G/N. Now $T_{a,b}$ and $T_{a,0}$ are in the same left coset of N in G, so in G/N we are getting an element by identifying $T_{a,b}$ with $T_{a,0}$. The latter element just depends on a. Moreover, the $T_{a,b}$ multiply according to $T_{a,b}T_{c,d} = T_{ac,ad+b}$ and if we identify $T_{a,b}$ with $T_{a,0}$, $T_{c,d}$ with $T_{c,0}$, then their product, which is $T_{ac,ad+b}$, is identified with $T_{ac,0}$. So in G/N multiplication is like that of the group of nonzero real numbers under multiplication, and in some sense (which will be made more precise in the next section) G/N can be identified with this group of real numbers.

2. Let G be the group of real numbers under $+$ and let $\mathbb{Z}$ be the group of integers under $+$. Since G is abelian, $\mathbb{Z} \lhd G$, and so we can talk about $G/\mathbb{Z}$. What does $G/\mathbb{Z}$ really look like? *In forming $G/\mathbb{Z}$, we are identifying any two real numbers that differ by an integer.* So 0 is identified with $-1, -2, -3, \ldots$ and $1, 2, 3, \ldots$; $\frac{3}{2}$ is identified with $\frac{1}{2}, \frac{5}{2}, -\frac{1}{2}, -\frac{3}{2}, \ldots$. Every real number a thus has a mate, $\tilde{a}$, where $0 \leq \tilde{a} < 1$. So, in $G/\mathbb{Z}$, the whole real line has been compressed into the unit interval $[0, 1]$. But a little more is true, for we have also identified the end points of this unit interval. So we are bending the unit interval around so that its two end points touch and become one. What do we get this way? A circle, of course! So $G/\mathbb{Z}$ is like a circle, in a sense that can be made precise, and this circle is a group with an appropriate product.

3. Let G be the group of nonzero complex numbers and let $N = \{a \in G \mid |a| = 1\}$ which is the unit circle in the complex plane. Then N is a subgroup of G and is normal since G is abelian. In going to G/N we are declaring that any complex number of absolute value 1 will be identified with the real number 1. Now any $a \in G$, in its polar form, can be written as $a = r(\cos\theta + i \sin\theta)$, where $r = |a|$, and $|\cos\theta + i \sin\theta| = 1$. In identifying $\cos\theta + i \sin\theta$ with 1, we are identifying a with r. So in passing to G/N every element is being identified with a positive real number, and this identification jibes with the products in G and in the group of positive real numbers, since $|ab| = |a||b|$. So G/N is in a very real sense (no pun intended) the group of positive real numbers under multiplication.

PROBLEMS

1. If G is the group of all nonzero real numbers under multiplication and N is the subgroup of all positive real numbers, write out G/N by exhibiting the cosets of N in G, and construct the multiplication in G/N.
2. If G is the group of nonzero real numbers under multiplication and

$N = \{1, -1\}$, show how you can "identify" G/N as the group of all positive real numbers under multiplication. What are the cosets of N in G?

3. If G is a group and $N \lhd G$, show that if $\overline{M}$ is a subgroup of G/N and $M = \{a \in G \mid Na \in \overline{M}\}$, then M is a subgroup of G, and $M \supset N$.
4. If $\overline{M}$ in Problem 3 is normal in G/N, show that the M defined is normal in G.
5. In Problem 3, show that M/N must equal $\overline{M}$.
6. Arguing as in the Example 2, where we identified $G/\mathbb{Z}$ as a circle, where G is the group of reals under $+$ and $\mathbb{Z}$ integers, consider the following: let $G = \{(a, b) \mid a, b \text{ real}\}$, where $+$ in G is defined by $(a, b) + (c, d) = (a + c, b + d)$ (so G is the plane), and let $N = \{(a, b) \in G \mid a, b \text{ are integers}\}$. Show that G/N can be identified as a torus (donut), and so we can define a product on the donut so that it becomes a group. Here, you may think of a torus as the Cartesian product of two circles.
7. If G is a cyclic group and N is a subgroup of G, show that G/N is a cyclic group.
8. If G is an abelian group and N is a subgroup of G, show that G/N is an abelian group.
9. Do Problems 7 and 8 by observing that G/N is a homomorphic image of G.
10. Let G be an abelian group of order $p_1^{a_1}p_2^{a_2} \cdots p_k^{a_k}$, where $p_1, p_2, \ldots, p_k$ are distinct prime numbers. Show that G has subgroups $S_1, S_2, \ldots, S_k$ of orders $p_1^{a_1}, p_2^{a_2}, \ldots, p_k^{a_k}$, respectively. (**Hint:** Use Cauchy's Theorem and pass to a factor group.) This result, which actually holds for all finite groups, is a famous result in group theory known as *Sylow's Theorem*. We prove it in Section 11.
11. If G is a group and $Z(G)$ the center of G, show that if $G/Z(G)$ is cyclic, then G is abelian.
12. If G is a group and $N \lhd G$ is such that G/N is abelian, prove that $aba^{-1}b^{-1} \in N$ for all $a, b \in G$.
13. If G is a group and $N \lhd G$ is such that

$$aba^{-1}b^{-1} \in N$$

for all $a, b \in G$, prove that G/N is abelian.

14. If G is an abelian group of order $p_1 p_2 \cdots p_k$, where $p_1, p_2, \ldots, p_k$ are distinct primes, prove that G is cyclic. (See Problem 15.)

15. If G is an abelian group and if G has an element of order m and one of order n, where m and n are relatively prime, prove that G has an element of order mn.

16. Let G be an abelian group of order $p^n m$, where p is a prime and $p \nmid m$. Let $P = \{a \in G \mid a^{p^k} = e \text{ for some } k \text{ depending on } a\}$. Prove that:

(a) P is a subgroup of G.

(b) G/P has no elements of order p.

(c) $|P| = p^n$.

17. Let G be an abelian group of order mn, where m and n are relatively prime. Let $M = \{a \in G \mid a^m = e\}$. Prove that:

(a) M is a subgroup of G.

(b) G/M has no element, x, other than the identity element, such that $x^m =$ unit element of G/M.

18. Let G be an abelian group (possibly infinite) and let the set $T = \{a \in G \mid a^m = e, m > 1 \text{ depending on } a\}$. Prove that:

(a) T is a subgroup of G.

(b) G/T has no element—other than its identity element—of finite order.

7. THE HOMOMORPHISM THEOREMS

Let G be a group and φ a homomorphism of G *onto* G'. If K is the kernel of φ, then K is a normal subgroup of G, hence we can form G/K. It is fairly natural to expect that there should be a very close relationship between G' and G/K. The *First Homomorphism Theorem*, which we are about to prove, spells out this relationship in exact detail.

But first let's look back at some of the examples of factor groups in Section 6 to see explicitly what the relationship mentioned above might be.

1. Let $G = \{T_{a,b} \mid a \neq 0, b \text{ real}\}$ and let G' be the group of nonzero reals under multiplication. From the product rule of these T's, namely $T_{a,b}T_{c,d} = T_{ac,ad+b}$, we determined that the mapping $\varphi: G \to G'$ defined by $\varphi(T_{a,b}) = a$ is a homomorphism of G onto G' with kernel $K = \{T_{1,b} \mid b \text{ real}\}$. On the other hand, in Example 1 of Section 6 we saw that $G/K = \{KT_{a,0} \mid a \neq 0 \text{ real}\}$. Since

$$(KT_{a,0})(KT_{x,0}) = KT_{ax,0}$$

the mapping of G/K onto G', which sends each $KT_{a,0}$ onto a, is readily seen to be an *isomorphism* of G/K onto G'. Therefore, $G/K \simeq G'$.

2. In Example 3, G was the group of nonzero complex numbers under multiplication and G' the group of all positive real numbers under multiplication.

Let $\varphi: G \to G'$ defined by $\varphi(a) = |a|$ for $a \in G$. Then, since $|ab| = |a|\,|b|$, φ is a homomorphism of G onto G' (can you see why it is onto?). Thus the kernel K of φ is precisely $K = \{a \in G \mid |a| = 1\}$. But we have already seen that if $|a| = 1$, then a is of the form $\cos\theta + i \sin\theta$. So the set $K = \{\cos\theta + i \sin\theta \mid 0 \leq \theta < 2\pi\}$. If a is any complex number, then $a = r(\cos\theta + i \sin\theta)$, where $r = |a|$, is the polar form of a. Thus $Ka = Kr(\cos\theta + i \sin\theta) = K(\cos\theta + i \sin\theta)r = Kr$, since $K(\cos\theta + i \sin\theta) = K$ because $\cos\theta + i \sin\theta \in K$. So G/K, whose elements are the cosets Ka, from this discussion, has all its elements of the form Kr, where $r > 0$. The mapping of G/K onto G' defined by sending Kr onto r then defines an *isomorphism* of G/K onto G'. So, here, too, $G/K \simeq G'$.

With this little experience behind us we are ready to make the jump the whole way, namely, to

Theorem 2.7.1 (First Homomorphism Theorem). Let φ be a homomorphism of G *onto* G' with kernel K. Then $G' \simeq G/K$, the isomorphism between these being effected by the map

$$\psi: G/K \to G'$$

defined by $\psi(Ka) = \varphi(a)$.

Proof. The best way to show that G/K and G' are isomorphic is to exhibit explicitly an isomorphism of G/K onto G'. The statement of the theorem suggests what such an isomorphism might be.

So define $\psi: G/K \to G'$ by $\psi(Ka) = \varphi(a)$ for $a \in G$. As usual, our first task is to show that ψ is well defined, that is, to show that if $Ka = Kb$, then $\psi(Ka) = \psi(Kb)$. This boils down to showing that if $Ka = Kb$, then $\varphi(a) = \varphi(b)$. But if $Ka = Kb$, then $a = kb$ for some $k \in K$, hence $\varphi(a) = \varphi(kb) = \varphi(k)\varphi(b)$. Since $k \in K$, the kernel of φ, then $\varphi(k) = e'$, the identity element of G', so we get $\varphi(a) = \varphi(b)$. This shows that the mapping ψ is well defined.

Because φ is onto G', given $x \in G'$, then $x = \varphi(a)$ for some $a \in G$, thus $x = \varphi(a) = \psi(Ka)$. This shows that ψ maps G/K onto G'.

Is ψ 1-1? Suppose that $\psi(Ka) = \psi(Kb)$; then $\varphi(a) = \psi(Ka) = \psi(Kb) = \varphi(b)$. Therefore, $e' = \varphi(a)\varphi(b)^{-1} = \varphi(a)\varphi(b^{-1}) = \varphi(ab^{-1})$. Because ab^{-1} is thus in the kernel of φ—which is K—we have $ab^{-1} \in K$. This implies that $Ka = Kb$. In this way ψ is seen to be 1-1.

Finally, is ψ a homomorphism? We check: $\psi((Ka)(Kb)) = \psi(Kab) = \varphi(ab) = \varphi(a)\varphi(b) = \psi(Ka)\psi(Kb)$, using that φ is a homomorphism and that $(Ka)(Kb) = Kab$. Consequently, ψ is a homomorphism of G/K onto G', and Theorem 2.7.1 is proved. □

Having talked about the *First* Homomorphism Theorem suggests that there are others. The next result, however, is an extension of the First Homomorphism Theorem, and is traditionally called the *Correspondence Theorem*. In the context of the theorem above, it exhibits a 1-1 correspondence between subgroups of G^1 and those subgroups of G that contain K.

Theorem 2.7.2 (Correspondence Theorem). Let the map $\varphi: G \to G'$ be a homomorphism of G onto G' with kernel K. If H' is a subgroup of G' and if

$$H = \{a \in G \mid \varphi(a) \in H'\},$$

then H is a subgroup of G, $H \supset K$, and $H/K \simeq H'$. Finally, if $H' \lhd G'$, then $H \lhd G$.

Proof. We first verify that the H above is a subgroup of G. It is not empty, since $e \in H$. If $a, b \in H$, then $\varphi(a), \varphi(b) \in H'$, hence $\varphi(ab) = \varphi(a)\varphi(b) \in H'$, since H' is a subgroup of G'; this puts ab in H, so H is closed. Further, if $a \in H$, then $\varphi(a) \in H'$, hence $\varphi(a^{-1}) = \varphi(a)^{-1}$ is in H', again since H' is a subgroup of G', whence $a^{-1} \in H$. Therefore, H is a subgroup of G.

Because $\varphi(K) = \{e'\} \subset H'$, where e' is the unit element of G', we have that $K \subset H$. Since $K \lhd G$ and $K \subset H$, it follows that $K \lhd H$. The mapping φ restricted to H defines a homomorphism of H onto H' with kernel K. By the First Homomorphism Theorem we get $H/K \simeq H'$.

Finally, if $H' \lhd G'$ and if $a \in G$, then $\varphi(a)^{-1}H'\varphi(a) \subset H'$, so $\varphi(a^{-1})H'\varphi(a) \subset H'$. This tells us that $\varphi(a^{-1}Ha) \subset H'$, so $a^{-1}Ha \subset H$. This proves the normality of H in G. □

It is worth noting that if K is any normal subgroup of G, and φ is the natural homomorphism of G onto G/K, then the theorem gives us a 1-1 correspondence between all subgroups H' of G/K and those subgroups of G that contain K. Moreover, this correspondence preserves normality in the sense that H' is normal in G/K if and only if H is normal in G. (See Problem 7, as well as the last conclusion of the theorem.)

We now state the *Second Homomorphism Theorem*, leaving its proof to the reader in Problem 5.

Theorem 2.7.3 (Second Homomorphism Theorem). Let H be a subgroup of a group G and N a normal subgroup of G. Then $HN = \{hn \mid h \in H, n \in N\}$ is a subgroup of G, $H \cap N$ is a normal subgroup of H, and $H/(H \cap N) \simeq (HN)/N$.

Finally, we go on to the *Third Homomorphism Theorem*, which tells us a little more about the relationship between N and N' when $N' \lhd G'$.

Theorem 2.7.4 (Third Homomorphism Theorem). If the map $\varphi: G \to G'$ is a homomorphism of G onto G' with kernel K then, if $N' \lhd G'$ and $N = \{a \in G \mid \varphi(a) \in N'\}$, we conclude that $G/N \simeq G'/N'$. Equivalently, $G/N \simeq (G/K)/(N/K)$.

Proof. Define the mapping $\psi: G \to G'/N'$ by $\psi(a) = N'\varphi(a)$ for every $a \in G$. Since φ is onto G' and every element of G'/N' is a coset of the form $N'x'$, and $x' = \varphi(x)$ for some $x \in G$, we see that ψ maps G onto G'/N'.

Furthermore, ψ is a homomorphism of G onto G'/N', for $\psi(ab) = N'\varphi(ab) = N'\varphi(a)\varphi(b) = (N'\varphi(a))(N'\varphi(b)) = \psi(a)\psi(b)$, since $N' \lhd G'$. What is the kernel, M, of ψ? If $a \in M$, then $\psi(a)$ is the unit element of G'/N', that is, $\psi(a) = N'$. On the other hand, by the definition of ψ, $\psi(a) = N'\varphi(a)$. Because $N'\varphi(a) = N'$ we must have $\varphi(a) \in N'$; but this puts a in N, by the very definition of N. Thus $M \subset N$. That $N \subset M$ is easy and is left to the reader. Therefore, $M = N$, so ψ is a homomorphism of G onto G'/N' with kernel N, whence, by the First Homomorphism Theorem, $G/N \simeq G'/N'$.

Finally, again by Theorems 2.7.1 and 2.7.2, $G' \simeq G/K$, $N' \simeq N/K$, which leads us to $G/N \simeq G'/N' \simeq (G/K)/(N/K)$. □

This last equality is highly suggestive; we are sort of "canceling out" the K in the numerator and denominator.

PROBLEMS

1. Show that $M \supset N$ in the proof of Theorem 2.7.3.
2. Let G be the group of all real-valued functions on the unit interval $[0, 1]$, where we define, for $f, g \in G$, addition by $(f + g)(x) = f(x) + g(x)$ for every $x \in [0, 1]$. If $N = \{f \in G \mid f(\frac{1}{4}) = 0\}$, prove that $G/N \simeq$ real numbers under $+$.
3. Let G be the group of nonzero real numbers under multiplication and let $N = \{1, -1\}$. Prove that $G/N \simeq$ positive real numbers under multiplication.
4. If G_1, G_2 are two groups and $G = G_1 \times G_2 = \{(a, b) \mid a \in G_1, b \in G_2\}$, where we define $(a, b)(c, d) = (ac, bd)$, show that:
   (a) $N = \{(a, e_2) \mid a \in G_1\}$, where e_2 is the unit element of G_2, is a normal subgroup of G.
   (b) $N \simeq G_1$.
   (c) $G/N \simeq G_2$.

5. Let G be a group, H a subgroup of G, and $N \lhd G$. Let the set $HN = \{hn \mid h \in H, n \in N\}$. Prove that:

(a) $H \cap N \lhd H$.

(b) HN is a subgroup of G.

(c) $N \subset HN$ and $N \lhd HN$.

(d) $(HN)/N \simeq H/(H \cap N)$.

***6.** If G is a group and $N \lhd G$, show that if $a \in G$ has finite order $o(a)$, then Na in G/N has finite order m, where $m \mid o(a)$. (Prove this by using the homomorphism of G onto G/N.)

7. If φ is a homomorphism of G onto G' and $N \lhd G$, show that $\varphi(N) \lhd G'$.

8. CAUCHY'S THEOREM

In Theorem 2.6.4—Cauchy's Theorem—we proved that if a prime p divides the order of a finite *abelian* group G, then G contains an element of order p. We did point out there that Cauchy's Theorem is true even if the group is not abelian. We shall give a very neat proof of this here; this proof is due to McKay.

We return for a moment to set theory, doing something that we mentioned in the problems in Section 4.

Let S be a set, $f \in A(S)$, and define a relation on S as follows: $s \sim t$ if $t = f^i(s)$ for some integer i (i can be positive, negative, or zero). We leave it to the reader as a problem that this does indeed define an equivalence relation on S. The equivalence class of s, $[s]$, is called the *orbit* of s under f. So S is the disjoint union of the orbits of its elements.

When f is of order p, p a prime, we can say something about the size of the orbits under f; those of the readers who solved Problem 34 of Section 4 already know the result. We prove it here to put it on the record officially.

[If $f^k(s) = s$, of course $f^{tk}(s) = s$ for every integer t. (Prove!)]

Lemma 2.8.1. If $f \in A(S)$ is of order p, p a prime, then the orbit of any element of S under f has 1 or p elements.

Proof. Let $s \in S$; if $f(s) = s$, then the orbit of s under f consists merely of s itself, so has one element. Suppose then that $f(s) \neq s$. Consider the elements $s, f(s), f^2(s), \ldots, f^{p-1}(s)$; we claim that these p elements are distinct and constitute the orbit of s under f. If not, then $f^i(s) = f^j(s)$ for some $0 \leq i < j \leq p - 1$, which gives us that $f^{j-i}(s) = s$. Let $m = j - i$; then $0 < m \leq p - 1$ and $f^m(s) = s$. But $f^p(s) = s$ and since $p \nmid m$, $ap + bm = 1$ for some integers a and b. Thus $f^1(s) = f^{ap+bm}(s) = f^{ap}(f^{bm}(s)) = f^{ap}(s) = s$,

since $f^m(s) = f^p(s) = s$. This contradicts that $f(s) \neq s$. Thus the orbit of s under f consists of $s, f(s), f^2(s), \ldots, f^{p-1}(s)$, so as p elements. □

We now give McKay's proof of Cauchy's Theorem.

Theorem 2.8.2 (Cauchy). If p is a prime and p divides the order of G, then G contains an element of order p.

Proof. If $p = 2$, the result amounts to Problem 18 in Section 1. Assume that $p \neq 2$. Let S be the set of all *ordered* p-tuples $(a_1, a_2, \ldots, a_{p-1}, a_p)$, where $a_1, a_2, \ldots, a_p$ are in G and where $a_1a_2 \cdots a_{p-1}a_p = e$. We claim that S has n^{p-1} elements where $n = |G|$. Why? We can choose $a_1, \ldots, a_{p-1}$ arbitrarily in G, and by putting $a_p = (a_1a_2 \cdots a_{p-1})^{-1}$, the p-tuple $(a_1, a_2, \ldots, a_{p-1}, a_p)$ then satisfies

$$a_1a_2 \cdots a_{p-1}a_p = a_1a_2 \cdots a_{p-1}(a_1a_2 \cdots a_{p-1})^{-1} = e,$$

so is in S. Thus S has n^{p-1} elements.

Note that if $a_1a_2 \cdots a_{p-1}a_p = e$, then $a_pa_1a_2 \cdots a_{p-1} = e$ (for if $xy = e$ in a group, then $yx = e$). So the mapping $f: S \to S$ defined by $f(a_1, \ldots, a_p) = (a_p, a_1, a_2, \ldots, a_{p-1})$ is in $A(S)$. Note that $f \neq i$, the identity map on S, and that $f^p = i$, so f is of order p.

If the orbit of s under f has one element, then $f(s) = s$. On the other hand, if $f(s) \neq s$, we know that the orbit of s under f consists precisely of p distinct elements; this we have by Lemma 2.8.1. Now when is $f(s) \neq s$? We claim that $f(s) \neq s$ if and only if when $s = (a_1, a_2, \ldots, a_p)$, then for some $i \neq j$, $a_i \neq a_j$. (We leave this to the reader.) So $f(s) = s$ if and only if $s = (a, a, \ldots, a)$ for some $a \in G$.

Let m be the number of $s \in S$ such that $f(s) = s$; since for $s = (e, e, \ldots, e)$, $f(s) = s$, we know that $m \geq 1$. On the other hand, if $f(s) \neq s$, the orbit of s consists of p elements, and these orbits are disjoint, for they are equivalence classes. If there are k such orbits where $f(s) \neq s$, we get that $n^{p-1} = m + kp$, for we have accounted this way for every element of S.

But $p \mid n$ by assumption and $p \mid (kp)$. So we must have $p \mid m$, since $m = n^{p-1} - kp$. Because $m \neq 0$ and $p \mid m$, we get that $m > 1$. But this says that there is an $s = (a, a, \ldots, a) \neq (e, e, \ldots, e)$ in S; from the definition of S this implies that $a^p = e$. Since $a \neq e$, a is the required element of order p. □

Note that the proof tells us that the number of solutions in G of $x^p = e$ is a positive multiple of p.

We strongly urge the reader who feels uncomfortable with the proof just given to carry out its details for $p = 3$. In this case the action of f on S becomes clear and our assertions about this action can be checked explicitly.

Cauchy's Theorem has many consequences. We shall present one of these, in which we determine completely the nature of certain groups of order pq, where p and q are distinct primes. Other consequences will be found in the problem set to follow, and in later material on groups.

Lemma 2.8.3. Let G be a group of order pq, where p, q are primes and $p > q$. If $a \in G$ is of order p and A is the subgroup of G generated by a, then $A \lhd G$.

Proof. We claim that A is the *only* subgroup of G of order p. Suppose that B is another subgroup of order p. Consider the set $AB = \{xy \mid x \in A, y \in B\}$; we claim that AB has p^2 distinct elements. For suppose that $xy = uv$ where $x, u \in A$, $y, v \in B$; then $u^{-1}x = vy^{-1}$. But $u^{-1}x \in A$, $vy^{-1} \in B$, and since $u^{-1}x = vy^{-1}$, we have $u^{-1}x \in A \cap B$. Since $B \neq A$ and $A \cap B$ is a subgroup of A and A is of prime order, we are forced to conclude that $A \cap B = (e)$ and so $u^{-1}x = e$, that is, $u = x$. Similarly, $v = y$. Thus the number of distinct elements in AB is p^2. But all these elements are in G, which has only $pq < p^2$ elements (since $p > q$). With this contradiction we see that $B = A$ and A is the only subgroup of order p in G. But if $x \in G$, $B = x^{-1}Ax$ is a subgroup of G of order p, in consequence of which we conclude that $x^{-1}Ax = A$; hence $A \lhd G$. □

Corollary. If G, a are as in Lemma 2.8.3 and if $x \in G$, then $x^{-1}ax = a^i$, where $0 < i < p$, for some i (depending on x).

Proof. Since $e \neq a \in A$ and $x^{-1}Ax = A$, $x^{-1}ax \in A$. But every element of A is of the form a^i, $0 \leq i < p$, and $x^{-1}ax \neq e$. In consquence, $x^{-1}ax = a^i$, where $0 < i < p$. □

We now prove a result of a different flavor.

Lemma 2.8.4. If $a \in G$ is of order m and $b \in G$ is of order n, where m and n are relatively prime and $ab = ba$, then $c = ab$ is of order mn.

Proof. Suppose that A is the subgroup generated by a and B that generated by b. Because $|A| = m$ and $|B| = n$ and $(m, n) = 1$, we get $A \cap B = (e)$, which follows from Lagrange's Theorem, for $|A \cap B| \mid n$ and $|A \cap B| \mid m$.

Suppose that $c^i = e$, where $i > 0$; thus $(ab)^i = e$. Since $ab = ba$, $e = (ab)^i = a^i b^i$; this tells us that $a^i = b^{-i} \in A \cap B = (e)$. So $a^i = e$, whence $m \mid i$, and $b^i = e$, whence $n \mid i$. Because $(m, n) = 1$ and m and n both divide i, mn divides i. So $i \geq mn$. Since $(ab)^{mn} = a^{mn}b^{mn} = e$, we see that mn is the smallest positive integer i such that $(ab)^i = e$. This says that ab is of order mn, as claimed in the lemma. □

Before considering the more general case of groups of order pq, let's look at a special case, namely, a group G of order 15. By Cauchy's Theorem, G has elements b of order 3 and a of order 5. By the Corollary to Lemma 2.8.3, $b^{-1}ab = a^i$, where $0 < i < 5$. Thus

$$b^{-2}ab^2 = b^{-1}(b^{-1}ab)b = b^{-1}a^ib = (b^{-1}ab)^i = (a^i)^i = a^{i^2}$$

and similarly, $b^{-3}ab^3 = a^{i^3}$. But $b^3 = e$, so we get $a^{i^3} = a$, whence $a^{i^3-1} = e$. Since a is of order 5, 5 must divide $i^3 - 1$, that is, $i^3 \equiv 1(5)$. However, by Fermat's Theorem (Corollary to Theorem 2.4.8), $i^4 \equiv 1(5)$. These two equations for i tell us that $i \equiv 1(5)$, so, since $0 < i < 5$, $i = 1$. In short, $b^{-1}ab = a^i = a$, which means that $ab = ba$. Since a is of order 5 and b of order 3, by Lemma 2.8.4, $c = ab$ is of order 15. This means that the 15 powers $e = c^0, c, c^2, \ldots, c^{14}$ are distinct, so must sweep out all of G. In a word, *G must be cyclic.*

The argument given for 15 could have been made shorter, but the form in which we did it is the exact prototype for the proof of the more general

Theorem 2.8.5. Let G be a group of order pq, where p, q are primes and $p > q$. If $q \nmid p - 1$, then G must be cyclic.

Proof. By Cauchy's Theorem, G has an element a of order p and an element b of order q. By the Corollary to Lemma 2.8.3, $b^{-1}ab = a^i$ for some i with $0 < i < p$. Thus $b^{-r}ab^r = a^{i^r}$ for all $r \geq 0$ (Prove!), and so $b^{-q}ab^q = a^{i^q}$. But $b^q = e$; therefore, $a^{i^q} = a$ and so $a^{i^q-1} = e$. Because a is of order p, we conclude that $p \mid i^q - 1$, which is to say, $i^q \equiv 1(p)$. However, by Fermat's Theorem, $i^{p-1} \equiv 1(p)$. Since $q \nmid p - 1$, we conclude that $i \equiv 1(p)$, and since $0 < i < p$, $i = 1$ follows. Therefore, $b^{-1}ab = a^i = a$, hence $ab = ba$. By Lemma 2.8.4, $c = ab$ has order pq, so the powers of c sweep out all of G. Thus G is cyclic, and the theorem is proved. □

PROBLEMS

Middle-Level Problems

1. In the proof of Theorem 2.8.2, show that if some two entries in $s = (a_1, a_2, \ldots, a_p)$ are different, then $f(s) \neq s$, and the orbit of s under f has p elements.
2. Prove that a group of order 35 is cyclic.
3. Using the result of Problem 40 of Section 5, give another proof of Lemma 2.8.3. (**Hint:** Use for H a subgroup of order p.)
4. Construct a nonabelian group of order 21. (**Hint:** Assume that $a^3 = e$,

$b^7 = e$ and find some i such that $a^{-1}ba = a^i \neq a$, which is consistent with the relations $a^3 = b^7 = e$.)

5. Let G be a group of order $p^n m$, where p is prime and $p \nmid m$. Suppose that G has a normal subgroup P of order p^n. Prove that $\theta(P) = P$ for every automorphism θ of G.
6. Let G be a finite group with subgroups A, B such that $|A| > \sqrt{|G|}$ and $|B| > \sqrt{|G|}$. Prove that $A \cap B \neq (e)$.
7. If G is a group with subgroups A, B of orders m, n, respectively, where m and n are relatively prime, prove that the subset of G, $AB = \{ab \mid a \in A, b \in B\}$, has mn distinct elements.
8. Prove that a group of order 99 has a nontrivial normal subgroup.
9. Prove that a group of order 42 has a nontrivial normal subgroup.
10. From the result of Problem 9, prove that a group of order 42 has a normal subgroup of order 21.

Harder Problems

11. If G is a group and A, B finite subgroups of G, prove that the set $AB = \{ab \mid a \in A, b \in B\}$ has $(|A|\,|B|)/|A \cap B|$ distinct elements.
12. Prove that any two nonabelian groups of order 21 are isomorphic. (See Problem 4.)

Very Hard Problems

13. Using the fact that any group of order 9 is abelian, prove that any group of order 99 is abelian.
14. Let $p > q$ be two primes such that $q \mid p - 1$. Prove that there exists a nonabelian group of order pq. (**Hint:** Use the result of Problem 40 of Section 4, namely that U_p is cyclic if p is a prime, and the idea needed to do Problem 4 above.)
15. Prove that if $p > q$ are two primes such that $q \mid p - 1$, then any two nonabelian groups of order pq are isomorphic.

9. DIRECT PRODUCTS

In several of the problems and examples that appeared earlier, we went through the following construction: If G_1, G_2 are two groups, then $G = G_1 \times G_2$ is the set of all ordered pairs (a, b), where $a \in G_1$ and $b \in G_2$ and

where the product was defined *component-wise* via $(a_1, b_1)(a_2, b_2) = (a_1a_2, b_1b_2)$, the products in each component being carried out in the respective groups G_1 and G_2. We should like to formalize this procedure here.

Definition. If $G_1, G_2, \ldots, G_n$ are n groups, then their (*external*) *direct product* $G_1 \times G_2 \times G_3 \times \cdots \times G_n$ is the set of all ordered n-tuples $(a_1, a_2, \ldots, a_n)$ where $a_i \in G_i$, for $i = 1, 2, \ldots, n$, and where the product in $G_1 \times G_2 \times \cdots \times G_n$ is defined component-wise, that is,

$$(a_1, a_2, \ldots, a_n)(b_1, b_2, \ldots, b_n) = (a_1b_1, a_2b_2, \ldots, a_nb_n).$$

That $G = G_1 \times G_2 \times \cdots \times G_n$ is a group is immediate, with $(e_1, e_2, \ldots, e_n)$ as its unit element, where e_i is the unit element of G_i, and where $(a_1, a_2, \ldots, a_n)^{-1} = (a_1^{-1}, a_2^{-1}, \ldots, a_n^{-1})$.

G is merely the Cartesian product of the groups $G_1, G_2, \ldots, G_n$ with a product defined in G by component-wise multiplication. We call it *external*, since the groups $G_1, G_2, \ldots, G_n$ are any groups, with no relation necessarily holding among them.

Consider the subsets $\overline{G}_i \subset G_1 \times G_2 \times \cdots \times G_n = G$, where

$$\overline{G}_i = \{(e_1, \ldots, e_{i-1}, a_i, e_{i+1}, \ldots, e_n) \mid a_i \in G_i\};$$

in other words, $\overline{G}_i$ consists of all n-tuples where in the ith component any element of G_i can occur and where every other component is the identity element. Clearly, $\overline{G}_i$ is a group and is isomorphic to G_i by the isomorphism $\pi_i: \overline{G}_i \to G_i$ defined by $\pi_i(e_1, e_2, \ldots, a_i, \ldots, e_n) = a_i$. Furthermore, not only is $\overline{G}_i$ a subgroup of G but $\overline{G}_i \lhd G$. (Prove!)

Given any element $a = (a_1, a_2, \ldots, a_n) \in G$, then

$$a = (a_1, e_2, \ldots, e_n)(e_1, a_2, e_3, \ldots, e_n) \cdots (e_1, e_2, \ldots, e_{n-1}, a_n);$$

that is, every $a \in G$ can be written as $a = \bar{a}_1\bar{a}_2 \cdots \bar{a}_n$, where each $\bar{a}_i \in \overline{G}_i$. Moreover, a can be written in this way in a unique manner, that is, if $a = \bar{a}_1\bar{a}_2 \cdots \bar{a}_n = \bar{b}_1\bar{b}_2 \cdots \bar{b}_n$, where the $\bar{a}_i \in \overline{G}_i$ and $\bar{b}_i \in \overline{G}_i$, then $\bar{a}_1 = \bar{b}_1, \ldots, \bar{a}_n = \bar{b}_n$. So G is built up from certain normal subgroups, the $\overline{G}_i$, as $G = \overline{G}_1\overline{G}_2 \cdots \overline{G}_n$ in such a way that every element $a \in G$ has a *unique* representation in the form $a = \bar{a}_1\bar{a}_2 \cdots \bar{a}_n$ with $\bar{a}_i \in \overline{G}_i$.

This motivates the following

Definition. The group G is said to be the (*internal*) *direct product* of its *normal* subgroups $N_1, N_2, \ldots, N_n$ if every $a \in G$ has a *unique* representation in the form $a = a_1a_2 \cdots a_n$, where each $a_i \in N_i$ for $i = 1, 2, \ldots, n$.

From what we have discussed above we have the

Lemma 2.9.1. If $G = G_1 \times G_2 \times \cdots \times G_n$ is the external direct product of $G_1, G_2, \ldots, G_n$, then G is the internal direct product of the normal subgroups $\overline{G}_1, \overline{G}_2, \ldots, \overline{G}_n$ defined above.

We want to go in the other direction, namely to prove that if G is the internal direct product of its normal subgroups $N_1, N_2, \ldots, N_n$, then G is isomorphic to $N_1 \times N_2 \times \cdots \times N_n$. To do so, we first get some preliminary results.

The result we are about to prove has already occurred as Problem 20, Section 5. For the sake of completeness we prove it here.

Lemma 2.9.2. Let G be a group, M, N normal subgroups of G such that $M \cap N = (e)$. Then, given $m \in M$ and $n \in N$, $mn = nm$.

Proof. Consider the element $a = mnm^{-1}n^{-1}$. Viewing a as bracketed one way, $a = (mnm^{-1})n^{-1}$; then, since $N \lhd G$ and $n \in N$, $mnm^{-1} \in N$, so $a = (mnm^{-1})n^{-1}$ is also in N. Now bracket a in the other way, $a = m(nm^{-1}n^{-1})$. Since $M \lhd G$ and $m^{-1} \in M$, we have $nm^{-1}n^{-1} \in M$ and so $a = m(nm^{-1}n^{-1}) \in M$. Thus $a \in M \cap N = (e)$, which is to say, $mnm^{-1}n^{-1} = e$. This gives us that $mn = nm$, as required. □

If G is the internal direct product of the normal subgroups $N_1, N_2, \ldots, N_n$, we claim that $N_i \cap N_j = (e)$ for $i \neq j$. For suppose that $a \in N_i \cap N_j$; then $a = e \cdot e \cdots eae \cdots e$, where the a occurs in the ith place. This gives us one representation of a in $G = N_1N_2 \cdots N_n$. On the other hand, $a = e \cdot e \cdots e \cdot a \cdot e \cdots e$, where the a occurs in the jth place, so a has the second representation as an element of $N_1N_2 \cdots N_n$. By the uniqueness of the representation, we get $a = e$, and so $N_i \cap N_j = (e)$.

Perhaps things would be clearer if we do it for $n = 2$. So suppose that $N_1 \lhd G$, $N_2 \lhd G$, and every element $a \in G$ has a unique representation as $a = a_1 \cdot a_2$, where $a_1 \in N_1$, $a_2 \in N_2$. Suppose that $a \in N_1 \cap N_2$; then $a = a \cdot e$ is a representation of $a = a_1 \cdot a_2$ with $a_1 = a \in N_1$, $a_2 = e \in N_2$. However $a = e \cdot a$, so $a = b_1 \cdot b_2$, where $b_1 = e \in N_1$, $b_2 = a \in N_2$. By the uniqueness of the representation we must have $a_1 = b_1$, that is, $a = e$. So $N_1 \cap N_2 = (e)$.

The argument given above for $N_1, \ldots, N_n$ is the same argument as that given for $n = 2$, but perhaps is less transparent. At any rate we have proved

Lemma 2.9.3. If G is the internal direct product of its normal subgroups $N_1, N_2, \ldots, N_n$, then, for $i \neq j$, $N_i \cap N_j = (e)$.

Corollary. If G is as in Lemma 2.9.3, then if $i \neq j$ and $a_i \in N_i$ and $a_j \in N_j$, we have $a_ia_j = a_ja_i$.

Proof. By Lemma 2.9.3, $N_i \cap N_j = (e)$ for $i \neq j$. Since the N's are normal in G, by Lemma 2.9.2 we have that any element in N_i *commutes* with any element in N_j, that is, $a_i a_j = a_j a_i$ for $a_i \in N_i$, $a_j \in N_j$. □

With these preliminaries out of the way we can now prove

Theorem 2.9.4. Let G be a group with normal subgroups $N_1, N_2, \ldots, N_n$. Then the mapping $\psi(a_1, a_2, \ldots, a_n) = a_1 a_2 \cdots a_n$ is an isomorphism from $N_1 \times N_2 \times \cdots \times N_n$ (external direct product) onto G if and only if G is the internal direct product of $N_1, N_2, \ldots, N_n$.

Proof. Suppose G is an internal direct product of $N_1, \ldots, N_n$. Since every element a in G has a representation $a = a_1 a_2 \cdots a_n$, with the $a_i \in N_i$, we have that the mapping ψ is onto. We assert that it is also 1-1. For if $\psi((a_1, a_2, \ldots, a_n)) = \psi((b_1, b_2, \ldots, b_n))$, then by the definition of ψ, $a_1 a_2 \cdots a_n = b_1 b_2 \cdots b_n$. By the uniqueness of the representation of an element in this form we deduce that $a_1 = b_1, a_2 = b_2, \ldots, a_n = b_n$. Hence ψ is 1-1.

All that remains is to show that ψ is a homomorphism. So, consider

$$\begin{aligned}\psi((a_1, a_2, \ldots, a_n)(b_1, b_2, \ldots, b_n)) &= \psi((a_1 b_1, a_2 b_2, \ldots, a_n b_n)) \\ &= (a_1 b_1)(a_2 b_2) \cdots (a_n b_n) \\ &= a_1 b_1 a_2 b_2 \cdots a_n b_n.\end{aligned}$$

Since $b_1 \in N_1$, it commutes with a_i, b_i for $i > 1$ by the Corollary to Lemma 2.9.3. So we can pull the b_1 across all the elements to the right of it to get $a_1 b_1 a_2 b_2 \cdots a_n b_n = a_1 a_2 b_2 a_3 b_3 \cdots a_n b_n b_1$. Now repeat this procedure with b_2, and so on, to get that $a_1 b_1 a_2 b_2 \cdots a_n b_n = (a_1 a_2 \cdots a_n)(b_1 b_2 \cdots b_n)$. Thus

$$\begin{aligned}\psi((a_1, a_2, \ldots, a_n)(b_1, b_2, \ldots, b_n)) &= a_1 b_1 a_2 b_2 \cdots a_n b_n \\ &= (a_1 a_2 \cdots a_n)(b_1 b_2 \cdots b_n) \\ &= \psi((a_1, a_2, \ldots, a_n))\psi((b_1 b_2, \ldots, b_n)).\end{aligned}$$

In other words, ψ is a homomorphism.

On the other hand, suppose that ψ is an isomorphism. Then the conclusion that G is the internal direct product of $N_1, N_2, \ldots, N_n$ easily follows from the fact that ψ is onto and 1-1.

With this the proof of Theorem 2.9.4 is complete. □

Corollary. Let G be a group with normal subgroups N_1, N_2. Then G is the internal direct product of N_1 and N_2 if and only if $G = N_1 N_2$ and $N_1 \cap N_2 = (e)$.

Proof. This follows easily from the fact that $\psi: N_1 \times N_2 \to G$, which is given by $\psi(a_1, a_2) = a_1 a_2$, is an isomorphism if and only if $N_1 N_2 = G$ and $N_1 \cap N_2 = (e)$. □

In view of the result of Theorem 2.9.4 and its corollary, we drop the adjectives "internal" and "external" and merely speak about the "direct product." When notation $G = N_1 \times N_2$ is used it should be clear from context whether it stands for the internal or external direct product.

The objective is often to show that a given group is the direct product of certain normal subgroups. If one can do this, the structure of the group can be completely determined if we happen to know those of the normal subgroups.

PROBLEMS

1. If G_1 and G_2 are groups, prove that $G_1 \times G_2 \simeq G_2 \times G_1$.

2. If G_1 and G_2 are cyclic groups of orders m and n, respectively, prove that $G_1 \times G_2$ is cyclic if and only if m and n are relatively prime.

3. Let G be a group, $A = G \times G$. In A let $T = \{(g, g) \mid g \in G\}$.
(a) Prove that $T \simeq G$.
(b) Prove that $T \lhd A$ if and only if G is abelian.

4. Let G be an abelian group of order $p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$, where $p_1, p_2, \ldots, p_k$ are distinct primes and $m_1 > 0$, $m_2 > 0, \ldots, m_k > 0$. By Problem 10 of Section 6, for each i, G has a subgroup P_i of order $p_i^{m_i}$. Show that $G \simeq P_1 \times P_2 \times \cdots \times P_k$.

5. Let G be a finite group, $N_1, N_2, \ldots, N_k$ normal subgroups of G such that $G = N_1 N_2 \cdots N_k$ and $|G| = |N_1|\,|N_2| \cdots |N_k|$. Prove that G is the direct product of $N_1, N_2, \ldots, N_k$.

6. Let G be a group, $N_1, N_2, \ldots, N_k$ normal subgroups of G such that:
1. $G = N_1 N_2 \cdots N_k$.
2. For each i, $N_i \cap (N_1 N_2 \cdots N_{i-1} N_{i+1} \cdots N_k) = (e)$.

Prove that G is the direct product of $N_1, N_2, \ldots, N_k$.

10. FINITE ABELIAN GROUPS (OPTIONAL)

We have just finished discussing the idea of the direct product of groups. If we were to leave that topic at the point where we ended, it might seem like a nice little construction, but so what? To give some more substance to it,

we should prove at least one theorem which says that a group satisfying a certain condition is the direct product of some particularly easy groups. Fortunately, such a class of groups exists, the finite abelian groups. What we shall prove is that any finite abelian group is the direct product of cyclic groups. This reduces most questions about finite abelian groups to questions about cyclic groups, a reduction that often allows us to get complete answers to these questions.

The results on the structure of finite abelian groups are really special cases of some wider and deeper theorems. To consider these would be going too far afield, especially since the story for finite abelian groups is so important in its own right. The theorem we shall prove is called the *Fundamental Theorem on Finite Abelian Groups*, and rightfully so.

Before getting down to the actual details of the proof, we should like to give a quick sketch of how we shall go about proving the theorem.

Our first step will be to reduce the problem from any finite abelian group to one whose order is p^n, where p is a prime. This step will be fairly easy to carry out, and since the group will have order involving just one prime, the details of the proof will not be cluttered with elements whose orders are somewhat complicated.

So we shall focus on groups of order p^n. Let G be an abelian group of order p^n. We want to show that there exist cyclic subgroups of G, namely $A_1, A_2, \ldots, A_k$, such that every element $x \in G$ can be written as $x = b_1 b_2 \cdots b_k$, where each $b_i \in A_i$, in a unique way. Otherwise put, since each A_i is cyclic and generated by a_i, say, we want to show that $x = a_1^{m_1} a_2^{m_2} \cdots a_k^{m_k}$, where the elements $a_i^{m_i}$ are unique.

A difficulty appears right away, for there is not just one choice for these elements $a_1, \ldots, a_k$. For instance, if G is the abelian group of order 4 with elements e, a, b, ab, where $a^2 = b^2 = e$ and $ab = ba$, then we can see that if A, B, C are the cyclic subgroups generated by a, b, and ab, respectively, then $G = A \times B = A \times C = B \times C$. So there is a lack of uniqueness in the choice of the a_i. How to get around this?

What we need is a mechanism for picking a_1 and which, when applied after we have picked a_1, will allow us to pick a_2, and so on. What should this mechanism be? Our control on the elements of G lies only in specifying their orders. It is the order of the element—when properly used—that will give us the means to prove the theorem.

Suppose that $G = A_1 \times A_2 \times \cdots \times A_k$, where $|G| = p^n$ and the A's have been numbered, so that $|A_i| = p^{n_i}$ and $n_1 \geq n_2 \geq \cdots \geq n_k$, and each A_i is cyclic generated by a_i. If this were so and $x = a_1^{m_1} \cdots a_k^{m_k}$, then

$$x^{p^{n_1}} = (a_1^{m_1} \cdots a_k^{m_k})^{p^{n_1}} = a_1^{m_1 p^{n_1}} a_2^{m_2 p^{n_1}} \cdots a_k^{m_k p^{n_1}}$$

because $n_1 \geq n_i$, $p^{n_i} | p^{n_1}$, so since every $a_i^{m_i p^{n_i}} = e$, thus $x^{p^{n_1}} = e$. *In other words, a_1 should be an element of G whose order is as large as it can possibly be.* Fine, we can now pick a_1. What do we do for a_2? If $\overline{G} = G/A_1$, then to get the first element needed to represent $\overline{G}$ as a direct product of cyclic groups, we should pick an element in $\overline{G}$ whose order is maximal. What does this translate into in G itself? We want an element a_2 such that *a_2 requires as high a power as possible to fall into A_1*. So that will be the road to the selection of the second element. However, if we pick an element a_2 with this property, it may not do the trick; we may have to adapt it so that it will. The doing of all this is the technical part of the argument and does go through. Then one repeats it appropriately to find an element a_3, and so on.

This is the procedure we shall be going through to prove the theorem. But to smooth out these successive choices of $a_1, a_2, \ldots$, we shall use an induction argument and some subsidiary preliminary results.

With this sketch as guide we hope the proof of the theorem will make sense to the reader. One should not confuse the basic idea in the proof—which is quite reasonable—with the technical details, which may cloud the issue. So we now begin to fill in the details of the sketch of the proof that we outlined above.

Lemma 2.10.1. Let G be a finite abelian group of order mn, where m and n are relatively prime. If $M = \{x \in G \mid x^m = e\}$ and $N = \{x \in G \mid x^n = e\}$, then $G = M \times N$. Moreover, if neither m nor n is 1, then $M \neq (e)$ and $N \neq (e)$.

Proof. The sets M and N defined in the assertion above are quickly seen to be subgroups of G. Moreover, if $m \neq 1$, then by Cauchy's Theorem (Theorem 2.6.4) we readily obtain $M \neq (e)$, and similarly if $n \neq 1$, that $N \neq (e)$. Furthermore, since $M \cap N$ is a subgroup of both M and N, by Lagrange's Theorem, $|M \cap N|$ divides $|M| = m$ and $|N| = n$. Because m and n are relatively prime, we obtain $|M \cap N| = 1$, hence $M \cap N = (e)$.

To finish the proof, we need to show that $G = MN$ and $G = M \times N$. Since m and n are relatively prime, there exist integers r and s such that $rm + sn = 1$. If $a \in G$, then $a = a^1 = a^{sn+rm} = a^{sn}a^{rm}$; since $(a^{sn})^m = a^{snm} = e$, we have that $a^{sn} \in M$. Similarly, $a^{rm} \in N$. Thus $a = a^{sn}a^{rm}$ is in MN. In this way $G = MN$. It now follows from Corollary to Theorem 2.9.4 that $G = M \times N$. □

An immediate consequence is the

Corollary. Let G be a finite abelian group and let p be a prime such that p divides $|G|$. Then $G = P \times T$ for some subgroups P and T, where $|P| = p^m$, $m > 0$, and $|T|$ is not divisible by p.

Proof. Let $P = \{x \in G \mid x^{p^s} = e \text{ for some } s\}$ and let the subset $T = \{x \in G \mid x^t = e \text{ for } t \text{ relatively prime to } p\}$. By Lemma 2.10.1, $G = P \times T$ and $P \neq (e)$. Since every element in P has order a power of p, $|P|$ is not divisible by any other prime (by Cauchy's Theorem), so $|P| = p^m$ for some m.

It is easy to see that $p \nmid |T|$ by making use of Lagrange's Theorem. Thus we really have that P is not merely some subgroup of G but is what is called a p-Sylow subgroup of G. (See Section 11). □

We now come to the key step in the proof of the theorem we seek. The proof is a little difficult, but once we have this result the rest will be easy.

Theorem 2.10.2. Let G be an abelian group of order p^n, p a prime, and let $a \in G$ have maximal order of all the elements in G. Then $G = A \times Q$, where A is the cyclic subgroup generated by a.

Proof. We proceed by induction on n. If $n = 1$, then $|G| = p$ and G is already a cyclic group generated by any $a \neq e$ in G.

We suppose the theorem to be true for all $m < n$. We first show that the theorem is correct if there exists an element $b \in G$ such that $b \notin A = (a)$ and $b^p = e$. Let $B = (b)$, the subgroup of G generated by b; thus $A \cap B = (e)$ (see Problem 1).

Let $\overline{G} = G/B$; by assumption $B \neq (e)$, hence $|\overline{G}| < |G|$. In $\overline{G}$, what is the order of $\bar{a} = Ba$? We claim that $o(\bar{a}) = o(a)$. To begin with, we know that $o(\bar{a}) \mid o(a)$ (see Problem 6 of Section 2.7). On the other hand, $\bar{a}^{o(\bar{a})} = \bar{e}$, so $a^{o(\bar{a})} \in B$. Since $a^{o(\bar{a})} \in A$, we see that $a^{o(\bar{a})} \in A \cap B = (e)$, whence $a^{o(\bar{a})} = e$. This tells us that $o(a) \mid o(\bar{a})$. Hence $o(a) = o(\bar{a})$.

Since $\bar{a}$ is an element of maximal order in $\overline{G}$, by the induction we know that $\overline{G} = (\bar{a}) \times T$ for some subgroup T of $\overline{G}$. By the Correspondence Theorem we also know that $T = Q/B$ for some subgroup Q of G. We claim that G is the internal direct product $A \times Q$. That $G = AQ$ is left to the reader. It remains to show that $A \cap Q = (e)$. Let $a^i \in A \cap Q$. Then $\bar{a}^i \in Q/B = T$, and since $(\bar{a}) \cap T = (\bar{e})$, we have that $\bar{a}^i = \bar{e}$. But since $o(a) = o(\bar{a})$, this implies $a^i = e$. Therefore, $A \cap Q = (e)$ and we obtain that $G = A \times Q$.

Suppose, then, that there is *no* element b in G, b not in A, such that $b^p = e$. We claim that this forces $G = A = (a)$, in which case G is a cyclic group. Suppose that $G \neq A$ and let $x \in G$, $x \notin A$ have smallest possible order. Because $o(x^p) < o(x)$, we have, by our choice of x, that $x^p \in A$, hence $x^p = a^i$ for some i.

We claim that $p \mid i$. Let $o(a) = p^s$, and note that the maximality

of the order of a implies that $x^{p^s} = e$. But $x^{p^s} = (x^p)^{p^{s-1}} = (a^i)^{p^{s-1}} = e$. Since $o(a) = p^s$, we have $p \mid i$.

Thus $x^p = a^i$, where $p \mid i$. Let $y = a^{-i/p} \cdot x$. Then $y^p = a^{-i}x^p = a^{-i}a^i = e$. Moreover, $y \notin (a) = A$, because $x \notin A$. But this puts us back in the situation discussed above, where there exists a $b \in G$, $b \notin A$ such that $b^p = e$; in that case we saw that the theorem was correct. So we must have $G = (a)$, and G is a cyclic group. This finishes the induction and proves the theorem. □

We are now able to prove the very basic and important

Theorem 2.10.3 (Fundamental Theorem on Finite Abelian Groups). A finite abelian group is the direct product of cyclic groups.

Proof. Let G be a finite abelian group and p a prime that divides $|G|$. By the Corollary to Lemma 2.10.1, $G = P \times T$, where $|P| = p^n$. By Theorem 2.10.2, $P = A_1 \times A_2 \times \cdots \times A_k$, where the A_i are cyclic subgroups of P. Arguing by induction on $|G|$, we may thus assume that $T = T_1 \times T_2 \times \cdots \times T_q$, where the T_i are cyclic subgroups of T. Thus

$$\begin{aligned} G &= (A_1 \times A_2 \times \cdots \times A_k) \times (T_1 \times T_2 \times \cdots \times T_q) \\ &= A_1 \times A_2 \times \cdots \times A_k \times T_1 \times T_2 \times \cdots \times T_q. \end{aligned}$$

This very important theorem is now proved. □

We return to abelian groups G of order p^n. We now have at hand that $G = A_1 \times A_2 \times \cdots \times A_k$, where the A_i are cyclic groups of order p^{n_i}. We can arrange the numbering so that $n_1 \geq n_2 \geq \cdots \geq n_k$. Also, $|G| = |A_1 \times A_2 \times \cdots \times A_k| = |A_1|\,|A_2| \cdots |A_k|$, which gives us that

$$p^n = p^{n_1}p^{n_2} \cdots p^{n_k} = p^{n_1+n_2+\cdots+n_k},$$

hence $n = n_1 + n_2 + \cdots + n_k$. Thus the integers $n_i \geq 0$ give us a *partition* of n. It can be shown that these integers $n_1, n_2, \ldots, n_k$—which are called the *invariants* of G—are *unique*. In other words, two abelian groups of order p^n are isomorphic if and only if they have the same invariants. Granted this, it follows that the number of nonisomorphic abelian groups of order p^n is equal to the number of partitions of n.

For example, if $n = 3$, it has the following three partitions: $3 = 3$, $3 = 2 + 1$, $3 = 1 + 1 + 1$, so there are three nonisomorphic abelian groups of order p^3 (independent of p). The groups corresponding to these partitions are a cyclic group of order p^3, the direct product of a cyclic group of order p^2 by one of order p, and the direct product of three cyclic groups of order p, respectively.

For $n = 4$ we see the partitions are $4 = 4$, $4 = 3 + 1$, $4 = 2 + 2$, $4 = 2 + 1 + 1$, $4 = 1 + 1 + 1 + 1$, which are five in number. Thus there are five nonisomorphic groups of order p^4. Can you describe them via the partitions of 4?

Given an abelian group of order $n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$, where the p_i are distinct primes and the a_i are all positive, then G is the direct product of its so-called p_i– Sylow subgroups (see, e.g., the Corollary to Lemma 2.10.1). For each prime p_i there are as many groups of order $p_i^{a_i}$ as there are partitions of a_i. So the number of nonisomorphic abelian groups of order $n = p_1^{a_1} \cdots p_k^{a_k}$ is $f(a_1)f(a_2) \cdots f(a_k)$, where $f(m)$ denotes the number of partitions of m. Thus we know how many nonisomorphic finite abelian groups there are for any given order.

For instance, how many nonisomorphic abelian groups are there of order 144? Since $144 = 2^4 3^2$, and there are five partitions of 4, two partitions of 2, there are 10 nonisomorphic abelian groups of order 144.

The material treated in this section has been hard, the path somewhat tortuous, and the effort to understand quite intense. To spare the reader too much further agony, we assign only three problems to this section.

PROBLEMS

1. Let A be a normal subgroup of a group G, and suppose that $b \in G$ is an element of prime order p, and that $b \notin A$. Show that $A \cap (b) = (e)$.

2. Let G be an abelian group of order p^n, p a prime, and let $a \in G$ have maximal order. Show that $x^{o(a)} = e$ for all $x \in G$.

3. Let G be a finite group, with $N \lhd G$ and $a \in G$. Prove that:

(a) The order of aN in G/N divides the order of a in G, that is, $o(aN) \mid o(a)$.

(b) If $(a) \cap N = (e)$, then $o(aN) = o(a)$.

11. CONJUGACY AND SYLOW'S THEOREM (OPTIONAL)

In discussing equivalence relations in Section 4 we mentioned, as an example of such a relation in a group G, the notion of *conjugacy*. Recall that the element b in G is said to be *conjugate* to $a \in G$ (or merely, a conjugate of a) if there exists an $x \in G$ such that $b = x^{-1}ax$. We showed in Section 4 that this defines an equivalence relation on G. The equivalence class of a, which we denote by $\mathrm{cl}(a)$, is called the *conjugacy class* of a.

For a finite group an immediate question presents itself: How large is $\mathrm{cl}(a)$? Of course, this depends strongly on the element a. For instance, if $a \in Z(G)$, the center of G, then $ax = xa$ for all $x \in G$, hence $x^{-1}ax = a$; in other words, the conjugacy class of a in this case consists merely of the element a itself. On the other hand, if $\mathrm{cl}(a)$ consists only of the element a, then $x^{-1}ax = a$ for all $x \in G$. This gives us that $xa = ax$ for all $x \in G$, hence $a \in Z(G)$. So $Z(G)$ is characterized as the set of those elements a in G whose conjugacy class has only one element, a itself.

For an abelian group G, since $G = Z(G)$, two elements are conjugate if and only if they are equal. So conjugacy is not an interesting relation for abelian groups; however, for nonabelian groups it is a highly interesting notion.

Given $a \in G$, $\mathrm{cl}(a)$ consists of all $x^{-1}ax$ as x runs over G. So to determine which are the distinct conjugates of a, we need to know when two conjugates of a coincide, which is the same as asking: When is $x^{-1}ax = y^{-1}ay$? In this case, transposing, we obtain $a(xy^{-1}) = (xy^{-1})a$; in other words, xy^{-1} must commute with a. This brings us to a concept introduced as Example 10 in Section 3, *that of the centralizer of a in G*. We repeat something we did there.

Definition. If $a \in G$, then $C(a)$, the *centralizer of a in G*, is defined by $C(a) = \{x \in G \mid xa = ax\}$.

When $C(a)$ arose in Section 3 we showed that it was a subgroup of G. We record this now more officially as

Lemma 2.11.1. For $a \in G$, $C(a)$ is a subgroup of G.

As we saw above, the two conjugates $x^{-1}ax$ and $y^{-1}ay$ of a are equal only if $xy^{-1} \in C(a)$, that is, only if x and y are in the same right coset of $C(a)$ in G. On the other hand, if x and y are in the same right coset of $C(a)$ in G, then $xy^{-1} \in C(a)$, hence $xy^{-1}a = axy^{-1}$. This yields that $x^{-1}ax = y^{-1}ay$. So x and y give rise to the same conjugate of a if and only if x and y are in the same right coset of $C(a)$ in G. *Thus there are as many conjugates of a in G as there are right cosets of* $C(a)$ *in* G. This is most interesting when G is a finite group, for in that case the number of right cosets of $C(a)$ in G is what we called the *index*, $i_G(C(a))$, of $C(a)$ in G, and is equal to $|G|/|C(a)|$.

We have proved

Theorem 2.11.2. Let G be a finite group and $a \in G$; then the number of distinct conjugates of a in G equals the index of $C(a)$ in G.

In other words, the number of elements in $\mathrm{cl}(a)$ equals $i_G(C(a)) = |G|/|C(a)|$.

This theorem, although it was relatively easy to prove, is very important and has many consequences. We shall see a few of these here.

One such consequence is a kind of bookkeeping result. Since conjugacy is an equivalence relation on G, G is the union of the disjoint conjugacy classes. Moreover, by Theorem 2.11.2, we know how many elements there are in each class. Putting all this information together, we get

Theorem 2.11.3 (The Class Equation). If G is a finite group, then

$$|G| = \sum_a i_G(C(a)) = \sum_a \frac{|G|}{|C(a)|},$$

where the sum runs over one a from each conjugacy class.

It is almost a sacred tradition among mathematicians to give, as the first application of the class equation, a particular theorem about groups of order p^n, where p is a prime. Not wanting to be accused of heresy, we follow this tradition and prove the pretty and important

Theorem 2.11.4. If G is a group of order p^n, where p is a prime, then $Z(G)$, the center of G, is not trivial (i.e., there exists an element $a \neq e$ in G such that $ax = xa$ for all $x \in G$).

Proof. We shall exploit the class equation to carry out the proof. Let $z = |Z(G)|$; as we pointed out previously, z is then the number of elements in G whose conjugacy class has only one element. Since $e \in Z(G)$, $z \geq 1$. For any element b outside $Z(G)$, its conjugacy class contains more than one element and $|C(b)| < |G|$. Moreover, since $|C(b)|$ divides $|G|$ by Lagrange's theorem, $|C(b)| = p^{n(b)}$, where $1 \leq n(b) < n$. We divide the pieces of the class equation into two parts: that coming from the center, and the rest. We get, this way,

$$p^n = |G| = z + \sum_{b \notin Z(G)} \frac{|G|}{|C(b)|} = z + \sum_{n(b)<n} \frac{p^n}{p^{n(b)}} = z + \sum_{n(b)<n} p^{n-n(b)}.$$

Clearly, p divides the left-hand side, p^n, and divides $\Sigma_{n(b)<n} p^{n-n(b)}$. The net result of this is that $p \mid z$, and since $z \geq 1$, we have that z is *at least* p. So since $z = |Z(G)|$, there must be an element $a \neq e$ in $Z(G)$, which proves the theorem. □

This last theorem has an interesting application, which some readers may have seen in solving Problem 45 of Section 5. This is

Theorem 2.11.5. If G is a group of order p^2, where p is a prime, then G is abelian.

Proof. By Theorem 2.11.4, $Z(G) \neq (e)$, so that there is an element, a, of order p in $Z(G)$. If $A = (a)$, the subgroup generated by a, then $A \subset Z(G)$, hence $A \subset C(x)$ for all $x \in G$. Given $x \in G$, $x \notin A$, then $C(x) \supset A$ and $x \in C(x)$; so $|C(x)| > p$, yet $|C(x)|$ must divide p^2. The net result of this is that $|C(x)| = p^2$, so $C(x) = G$, whence $x \in Z(G)$. Since every element of G is in the center of G, G must be abelian. □

In the problems to come we shall give many applications of the nature of groups of order p^n, where p is a prime. *The natural attack on virtually all these problems follows the lines of the argument we are about to give.* We choose one of a wide possible set of choices to illustrate this technique.

Theorem 2.11.6. If G is a group of order p^n, p a prime, then G contains a normal subgroup of order p^{n-1}.

Proof. We proceed by induction on n. If $n = 1$, then G is of order p and (e) is the required normal subgroup of order $p^{1-1} = p^0 = 1$.

Suppose that we know that for some k every group of order p^k has a normal subgroup of order p^{k-1}. Let G be of order p^{k+1}; by Theorem 2.11.4 there exists an element a of order p in $Z(G)$, the center of G. Thus the subgroup $A = (a)$ generated by a is of order p and is normal in G. Consider $\Gamma = G/A$; Γ is a group of order $|G|/|A| = p^{k+1}/p = p^k$ by Theorem 2.6.3. Since Γ has order p^k, we know that Γ has a normal subgroup M of order p^{k-1}. Since Γ is a homomorphic image of G, by the Correspondence Theorem (Theorem 2.7.2) there is a normal subgroup N in G, $N \supset A$, such that $N/A = M$. But then we have

$$p^{k-1} = |M| = |N/A| = \frac{|N|}{|A|},$$

that is, $p^{k-1} = |N|/p$, leading us to $|N| = p^k$. Thus N is our required normal subgroup in G of order p^k. This completes the induction and so proves the theorem. □

By far the most important application we make of the class equation is the proof of a far-reaching theorem due to Sylow, a Norwegian mathematician, who proved it in 1871. We already showed this theorem to be true for abelian groups. We shall now prove it for any finite group. It is impossible to overstate the importance of *Sylow's Theorem* in the study of finite groups. Without it the subject would not get off the ground.

Theorem 2.11.7 (Sylow's Theorem). Suppose that G is a group of order $p^n m$, where p is a prime and $p \nmid m$. Then G has a subgroup of order p^n.

Proof. If $n = 0$, this is trivial. We therefore assume that $n \geq 1$. Here, again, we proceed by induction on $|G|$, assuming the result to be true for all groups H such that $|H| < |G|$.

Suppose that the result is false for G. Then, by our induction hypothesis, p^n *cannot divide* $|H|$ for any subgroup H of G if $H \neq G$. *In particular, if* $a \notin Z(G)$, *then* $C(a) \neq G$, *hence* $p^n \nmid |C(a)|$. Thus p divides $|G|/|C(a)| = i_G(C(a))$ for $a \notin Z(G)$.

Write down the class equation for G following the lines of the argument in Theorem 2.11.4. If $z = |Z(G)|$, then $z \geq 1$ and

$$p^n m = |G| = z + \sum_{a \notin Z(G)} i_G(C(a)).$$

But $p \mid i_G(C(a))$ if $a \notin Z(G)$, so $p \mid \Sigma_{a \notin Z(G)} i_G(C(a))$. Since $p \mid p^n m$, we get $p \mid z$. By Cauchy's Theorem there is an element a of order p in $Z(G)$. If A is the subgroup generated by a, then $|A| = p$ and $A \lhd G$, since $a \in Z(G)$. Consider $\Gamma = G/A$; $|\Gamma| = |G|/|A| = p^n m/p = p^{n-1} m$. Since $|\Gamma| < |G|$, by our induction hypothesis Γ has a subgroup M of order p^{n-1}. However, by the Correspondence Theorem there is a subgroup P of G such that $P \supset A$ and $P/A = M$. Therefore, $|P| = |M|\,|A| = p^{n-1} p = p^n$ and P is the sought-after subgroup of G of order p^n, contradicting our assumption that G had no such subgroup. This completes the induction, and Sylow's Theorem is established. □

Actually, Sylow's Theorem consists of three parts, of which we only proved the first. The other two are (assuming $p^n m = |G|$, where $p \nmid m$):

1. Any two subgroups of order p^n in G are conjugate; that is, if $|P| = |Q| = p^n$ for subgroups P, Q of G, then for some $x \in G$, $Q = x^{-1} P x$.
2. The number of subgroups of order p^n in G is of the form $1 + kp$ and divides $|G|$.

Since these subgroups of order p^n pop up all over the place, they are called *p-Sylow subgroups* of G. An abelian group has one p-Sylow subgroup for every prime p dividing its order. This is far from true in the general case. For instance, if $G = S_3$, the symmetric group of degree 3, which has order $6 = 2 \cdot 3$, there are three 2-Sylow subgroups (of order 2) and one 3-Sylow subgroup (or order 3).

For those who want to see several proofs of that part of Sylow's Theorem which we proved above, and of the other two parts, they might look at the appropriate section of our book *Topics in Algebra*.

PROBLEMS

Easier Problems

1. In S_3, the symmetric group of degree 3, find all the conjugacy classes, and check the validity of the class equation by determining the orders of the centralizers of the elements of S_3.
2. Do Problem 1 for G the dihedral group of order 8.
3. If $a \in G$, show that $C(x^{-1}ax) = x^{-1}C(a)x$.
4. If φ is an automorphism of G, show that $C(\varphi(a)) = \varphi(C(a))$ for $a \in G$.
5. If $|G| = p^3$ and $|Z(G)| \geq p^2$, prove that G is abelian.
6. If P is a p-Sylow subgroup of G and $P \lhd G$, prove that P is the only p-Sylow subgroup of G.
7. If $P \lhd G$, P a p-Sylow subgroup of G, prove that $\varphi(P) = P$ for every automorphism φ of G.
8. Use the class equation to give a proof of Cauchy's Theorem.

   If H is a subgroup of G, let $N(H) = \{x \in G \mid x^{-1}Hx = H\}$. This *does not mean* that $xa = ax$ whenever $x \in N(H)$, $a \in H$. For instance, if $H \lhd G$, then $N(H) = G$, yet H need not be in the center of G.
9. Prove that $N(H)$ is a subgroup of G, $H \subset N(H)$ and in fact $H \lhd N(H)$.
10. Prove that $N(x^{-1}Hx) = x^{-1}N(H)x$.
11. If P is a p-Sylow subgroup of G, prove that P is a p-Sylow subgroup of $N(P)$ and is the only p-Sylow subgroup of $N(P)$.
12. If P is a p-Sylow subgroup and $a \in G$ is of order p^m for some m, show that if $a^{-1}Pa = P$ then $a \in P$.
13. Prove that if G is a finite group and H is a subgroup of G, then the number of distinct subgroups $x^{-1}Hx$ of G equals $i_G(N(H))$.
14. If P is a p-Sylow subgroup of G, show that the number of distinct $x^{-1}Px$ *cannot* be a multiple of p.
15. If $N \lhd G$, let $B(N) = \{x \in G \mid xa = ax \text{ for all } a \in N\}$. Prove that $B(N) \lhd G$.

Middle-Level Problems

16. Show that a group of order 36 has a normal subgroup of order 3 or 9. (**Hint:** See Problem 40 of Section 5.)
17. Show that a group of order 108 has a normal subgroup of order 9 or 27.
18. If P is a p-Sylow subgroup of G, show that $N(N(P)) = N(P)$.
19. If $|G| = p^n$, show that G has a subgroup of order p^m for all $1 \leq m \leq n$.

20. If p^m divides $|G|$, show that G has a subgroup of order p^m.

21. If $|G| = p^n$ and $H \neq G$ is a subgroup of G, show that $N(H) \supsetneq H$.

22. Show that any subgroup of order p^{n-1} in a group G of order p^n is normal in G.

Harder Problems

23. Let G be a group, H a subgroup of G. Define for $a, b \in G$, $a \sim b$ if $b = h^{-1}ah$ for some $h \in H$. Prove that

(a) this defines an equivalence relation on G.

(b) If $[a]$ is the equivalence class of a, show that if G is a finite group, then $[a]$ has m elements where m is the index of $H \cap C(a)$ in H.

24. If G is a group, H a subgroup of G, define a relation $B \sim A$ for subgroups A, B of G by the condition that $B = h^{-1}Ah$ for some $h \in H$.

(a) Prove that this defines an equivalence relation on the set of subgroups of G.

(b) If G is finite, show that the number of distinct subgroups equivalent to A equals the index of $N(A) \cap H$ in H.

25. If P is a p-Sylow subgroup of G, let S be the set of all p-Sylow subgroups of G. For $Q_1, Q_2 \in S$ define $Q_1 \sim Q_2$ if $Q_2 = a^{-1}Q_1a$ with $a \in P$. Prove, using this relation, that if $Q \neq P$, then the number of distinct $a^{-1}Qa$, with $a \in P$, is a multiple of p.

26. Using the result of Problem 25, show that the number of p-Sylow subgroups of G is of the form $1 + kp$. (This is the third part of Sylow's Theorem.)

27. Let P be a p-Sylow subgroup of G, and Q another one. Suppose that $Q \neq x^{-1}Px$ for any $x \in G$. Let S be the set of all $y^{-1}Qy$, as y runs over G. For $Q_1, Q_2 \in S$ define $Q_1 \sim Q_2$ if $Q_2 = a^{-1}Q_1a$, where $a \in P$.

(a) Show that this implies that the number of distinct $y^{-1}Qy$ is a multiple of p.

(b) Using the result of Problem 14, show that the result of Part (a) cannot hold.

(c) Prove from this that given any two p-Sylow subgroups P and Q of G, then $Q = x^{-1}Px$ for some $x \in G$.
(This is the second part of Sylow's Theorem.)

28. If H is a subgroup of G of order p^m show that H is contained in some p-Sylow subgroup of G.

29. If P is a p-Sylow subgroup of G and $a, b \in Z(P)$ are conjugate in G, prove that they are already conjugate in $N(P)$.

3

THE SYMMETRIC GROUP

Ziyad Asaad

1. PRELIMINARIES

Let us recall a theorem proved in Chapter 2 for abstract groups. This result, known as *Cayley's Theorem* (Theorem 2.5.1), asserts that any group G is isomorphic to a subgroup of $A(S)$, the set of 1-1 mappings of the set S onto itself, for some suitable S. In fact, in the proof we gave we used for S the group G itself viewed merely as a set.

Historically, groups arose this way first, long before the notion of an abstract group was defined. We find in the work of Lagrange, Abel, Galois, and others, results on groups of permutations proved in the late eighteenth and early nineteenth centuries. Yet it was not until the mid-nineteenth century that Cayley more or less introduced the abstract concept of a group.

Since the structure of isomorphic groups is the same, Cayley's Theorem points out a certain universal character for the groups $A(S)$. If we knew the structure of all subgroups of $A(S)$ for any set S, we would know the structure of all groups. This is much too much to expect. Nevertheless, one could try to exploit this embedding of an arbitrary group G isomorphically into some $A(S)$. This has the advantage of transforming G as an abstract system into something more concrete, namely a set of nice mappings of some set onto itself.

We shall not be concerned with the subgroups of $A(S)$ for an arbitrary set S. If S is infinite, $A(S)$ is a very wild and complicated object. Even if S is finite, the complete nature of $A(S)$ is virtually impossible to determine.

In this chapter we consider only $A(S)$ for S a finite set. Recall that if S has n elements, then we call $A(S)$ the *symmetric group of degree n*, and denote it by S_n. The elements of S_n are called *permutations*; we shall denote them by lowercase Greek letters.

Since we multiplied two elements $\sigma, \tau \in A(S)$ by the rule $(\sigma\tau)(s) = \sigma(\tau(s))$ this will have the effect that when we introduce the appropriate symbols to represent the elements of S_n, these symbols, or permutations, will multiply from *right to left.* If the readers look at some other book on algebra, they should make sure which way the permutations are being multiplied: right to left or left to right. Very often, algebraists multiply permutations from *left to right.* To be consistent with our definition of the composition of elements in S_n, we do it from right to left.

By Cayley's Theorem we know that if G is a finite group of order n, then G is isomorphic to a subgroup of S_n and S_n has $n!$ elements. Speaking loosely, we usually say that G is a subgroup of S_n. Since n is so much smaller than $n!$ for n even modestly large, our group occupies only a tiny little corner in S_n. It would be desirable to embed G in an S_n for n as small as possible. For certain classes of finite groups this is achievable in a particularly nice way.

Let S be a finite set having n elements; we might as well suppose that $S = \{x_1, x_2, \ldots, x_n\}$. Given the permutation $\sigma \in S_n = A(S)$, then $\sigma(x_k) \in S$ for $k = 1, 2, \ldots, n$, so $\sigma(x_k) = x_{i_k}$ for some i_k, $1 \leq i_k \leq n$. Because σ is 1-1, if $j \neq k$, then $x_{i_j} = \sigma(x_j) \neq \sigma(x_k) = x_{i_k}$. Therefore, the numbers $i_1, i_2, \ldots, i_n$ are merely the numbers $1, 2, \ldots, n$ shuffled about in some order.

Clearly, the action of σ on S is determined by what σ does to the subscript j of x_j, so the symbol "x" is really excess baggage and, as such, can be discarded. In short, we may assume that $S = \{1, 2, \ldots, n\}$.

Let's recall what is meant by the *product* of two elements of $A(S)$. If $\sigma, \tau \in A(S)$, then we defined $\sigma\tau$ by $(\sigma\tau)(s) = \sigma(\tau(s))$ for every $s \in S$. We showed in Section 4 of Chapter 1 that $A(S)$ satisfied four properties that we used later as the model to define the notion of an abstract group. Thus S_n, in particular, is a group relative to the product of mappings.

Our first need is some handy way of denoting a permutation, that is, an element σ in S_n. One clear way is to make a table of what σ does to each element of S. This might be called the *graph* of σ. We did this earlier, writing out σ, say $\sigma \in S_3$, in the fashion: $\sigma: x_1 \to x_2, x_2 \to x_3, x_3 \to x_1$. But this is cumbersome and space consuming. We certainly can make it more compact by dropping the x's and writing $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$. In this symbol the number in the second row is the image under σ of the number in the first row directly above it. There is nothing holy about 3 in all this; it works equally well for any n.

If $\sigma \in S_n$ and $\sigma(1) = i_1, \sigma(2) = i_2, \ldots, \sigma(n) = i_n$, we use the symbol $\begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}$ to represent σ and we write $\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}$. Note that it is not necessary to write the first row in the usual order $1 \quad 2 \quad \cdots \quad n$; any way we write the first row, as long as we carry the i_j's along accordingly, we still have σ. For instance, in the example in S_3 cited, $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 \\ 3 & 2 & 1 \end{pmatrix}$.

If we know $\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}$, what is σ^{-1}? It is easy, just flip the symbol for σ over to get $\sigma^{-1} = \begin{pmatrix} i_1 & i_2 & \cdots & i_n \\ 1 & 2 & \cdots & n \end{pmatrix}$. (Prove!) In our example $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$, $\sigma^{-1} = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$. The identity element—which we shall write as e—is merely $e = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix}$.

How does the product in S_n translate in terms of these symbols? Since $\sigma\tau$ means: "First apply τ and to the result of this apply σ," in forming the product of the symbols for σ and τ we look at the number k in the first row of τ and see what number i_k is directly below k in the second row of τ. We then look at the spot i_k in the first row of σ and see what is directly below it in the second row of σ. This is the image of k under $\sigma\tau$. We then run through $k = 1, 2, \ldots, n$ and get the symbol for $\sigma\tau$. We just do this visually.

We illustrate this with two permutations

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 5 & 4 \end{pmatrix} \text{ and } \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 1 & 2 \end{pmatrix}$$

in S_5. Then $\sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 4 & 2 & 3 \end{pmatrix}$.

Even the economy achieved this way is not enough. After all, the first row is always $1 \quad 2 \quad \cdots \quad n$, so we could dispense with it, and write $\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}$ as $(i_1, i_2, \ldots, i_n)$. This is fine, but in the next section we shall find a better and briefer way of representing permutations.

PROBLEMS

1. Find the products:

(a) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 5 & 2 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 1 \end{pmatrix}$.

(b) $\begin{pmatrix}1&2&3&4&5\\2&1&3&4&5\end{pmatrix}\begin{pmatrix}1&2&3&4&5\\3&2&1&4&5\end{pmatrix}$.

(c) $\begin{pmatrix}1&2&3&4&5\\4&1&3&2&5\end{pmatrix}^{-1}\begin{pmatrix}1&2&3&4&5\\2&1&3&4&5\end{pmatrix}\begin{pmatrix}1&2&3&4&5\\4&1&3&2&5\end{pmatrix}$.

2. Evaluate all the powers of each permutation σ (i.e., find σ^k for all k).

(a) $\begin{pmatrix}1&2&3&4&5&6\\2&3&4&5&6&1\end{pmatrix}$.

(b) $\begin{pmatrix}1&2&3&4&5&6&7\\2&1&3&4&6&5&7\end{pmatrix}$.

(c) $\begin{pmatrix}1&2&3&4&5&6\\6&4&5&2&1&3\end{pmatrix}$.

3. Prove that $\begin{pmatrix}1&2&\cdots&n\\i_1&i_2&\cdots&i_n\end{pmatrix}^{-1}=\begin{pmatrix}i_1&i_2&\cdots&i_n\\1&2&\cdots&n\end{pmatrix}$.

4. Find the order of each element in Problem 2.

5. Find the order of the products you obtained in Problem 1.

2. CYCLE DECOMPOSITION

We continue the process of simplifying the notation used to represent a given permutation. In doing so, we get something more than just a new symbol; we get a device to decompose any permutation as a product of particularly nice permutations.

Definition. Let $i_1, i_2, \ldots, i_k$ be k distinct integers in $S = \{1, 2, \ldots, n\}$. The symbol $(i_1 \quad i_2 \quad \cdots \quad i_k)$ will represent the permutation $\sigma \in S_n$, where $\sigma(i_1) = i_2$, $\sigma(i_2) = i_3, \ldots, \sigma(i_j) = i_{j+1}$ for $j < k$, $\sigma(i_k) = i_1$, and $\sigma(s) = s$ for any $s \in S$ if s is different from $i_1, i_2, \ldots, i_k$.

Thus, in S_7, the permutation $(1 \quad 3 \quad 5 \quad 4)$ is the permutation $\begin{pmatrix}1&2&3&4&5&6&7\\3&2&5&1&4&6&7\end{pmatrix}$. We call a permutation of the form $(i_1 \quad i_2 \quad \cdots \quad i_k)$ a *k-cycle*. For the special case $k = 2$, the permutation $(i_1 \quad i_2)$ is called a *transposition*. Note that if $\sigma = (i_1 \quad i_2 \quad \cdots \quad i_k)$, then σ is also $(i_k \quad i_1 \quad i_2 \quad \cdots \quad i_{k-1})$, $(i_{k-1} \quad i_k \quad i_1 \quad i_2 \quad \cdots \quad i_{k-2})$, and so on. (Prove!) For example,

$$(1 \quad 3 \quad 5 \quad 4) = (4 \quad 1 \quad 3 \quad 5) = (5 \quad 4 \quad 1 \quad 3) = (3 \quad 5 \quad 4 \quad 1).$$

Two cycles, say a k-cycle and an m-cycle, are said to be *disjoint cycles* if

they have no integer in common. Whence $(1 \quad 3 \quad 5)$ and $(4 \quad 2 \quad 6 \quad 7)$ in S_7 are disjoint cycles.

Given two disjoint cycles in S_n, we claim that they commute. We leave the proof of this to the reader, with the suggestion that if σ, τ are disjoint cycles, the reader should verify that $(\sigma\tau)(i) = (\tau\sigma)(i)$ for every $i \in S = \{1, 2, \ldots, n\}$. We state this result as

Lemma 3.2.1. If $\sigma, \tau \in S_n$ are disjoint cycles, then $\sigma\tau = \tau\sigma$.

Let's consider a particular k-cycle $\sigma = (1 \quad 2 \quad \cdots \quad k)$ in S_n. Clearly, $\sigma(1) = 2$ by the definition given above; how is 3 related to 1? Since $\sigma(2) = 3$, we have $\sigma^2(1) = \sigma(2) = 3$. Continuing, we see that $\sigma^j(1) = j + 1$ for $j \leq k - 1$, while $\sigma^k(1) = 1$. In fact, we see that $\sigma^k = e$, where e is the identity element in S_n.

There are two things to be concluded from the paragraph above.

1. The order of a k-cycle, as an element of S_n, is k. (Prove!)
2. If $\sigma = (i_1 \quad i_2 \quad \cdots \quad i_k)$ is a k-cycle, then the *orbit* of i_1 under σ (see Problem 27 in Section 4 of Chapter 1) is $\{i_1, i_2, \ldots, i_k\}$. So we can see that the k-cycle $\sigma = (i_1 \quad i_2 \quad \cdots \quad i_k)$ is

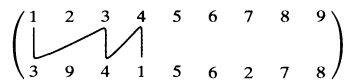
$$\sigma = (i_1 \quad \sigma(i_1) \quad \sigma^2(i_1) \quad \cdots \quad \sigma^{k-1}(i_1)).$$

Given *any* permutation τ in S_n, for $i \in \{1, 2, \ldots, n\}$, consider the *orbit* of i under τ. We have that this orbit is $\{i, \tau(i), \tau^2(i), \ldots, \tau^{s-1}(i)\}$, where $\tau^s(i) = i$ and s is the smallest positive integer with this property. Consider the s-cycle $(i \quad \tau(i) \quad \tau^2(i) \quad \cdots \quad \tau^{s-1}(i))$; we call it the *cycle of τ determined by i.*

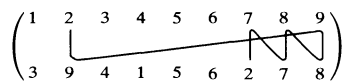
We take a specific example and find all its cycles. Let

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 9 & 4 & 1 & 5 & 6 & 2 & 7 & 8 \end{pmatrix};$$

what is the cycle of τ determined by 1? We claim that it is $(1 \quad 3 \quad 4)$. Why? τ takes 1 into 3, 3 into 4 and 4 into 1, and since $\tau(1) = 3$, $\tau^2(1) = \tau(3) = 4$, $\tau^3(1) = \tau(4) = 1$. We can get this visually by weaving through



with the thin path. What is the cycle of τ determined by 2? Weaving through



with the thin path, we see that the cycle of τ determined by 2 is $(2\ \ 9\ \ 8\ \ 7)$. The cycles of τ determined by 5 and 6 are (5) and (6), respectively, since 5 and 6 are left fixed by τ. So the cycles of τ are $(1\ \ 3\ \ 4)$, $(2\ \ 9\ \ 8\ \ 7)$, (5), and (6). Therefore we have that $\tau = (1\ \ 3\ \ 4)(2\ \ 9\ \ 8\ \ 7)$ (5)(6), where we view these cycles—as defined above—as permutations in S_9 because every integer in $S = \{1, 2, \ldots, 9\}$ appears in one and only one cycle, and the image of any i under τ is read off from the cycle in which it appears.

There is nothing special about the permutation τ above that made the argument we gave go through. The same argument would hold for *any* permutation in S_n for *any* n. We leave the formal writing down of the proof to the reader.

Theorem 3.2.2. Every permutation in S_n is the product of disjoint cycles.

In writing a permutation σ as a product of disjoint cycles, we omit all 1-cycles; that is, we ignore the i's such that $\sigma(i) = i$. Thus we write $\sigma = (1\ 2\ 3)(4\ 5)(6)(7)$ simply as $\sigma = (1\ 2\ 3)(4\ 5)$. In other words, writing σ as a product of k-cycles, with $k > 1$, we assume that σ leaves fixed any integer not present in any of the cycles. Thus in the group S_{11} the permutation $\tau = (1\ \ 5\ \ 6)(2\ \ 3\ \ 9\ \ 8\ \ 7)$ leaves fixed 4, 10, and 11.

Lemma 3.2.3. If τ in S_n is a k-cycle, then the order of τ is k; that is, $\tau^k = e$ and $\tau^j \neq e$ for $0 < j < k$.

Consider the permutation $\tau = (1\ \ 2)(3\ \ 4\ \ 5\ \ 6)(7\ \ 8\ \ 9)$ in S_9. What is its order? Since the disjoint cycles $(1\ \ 2)$, $(3\ \ 4\ \ 5\ \ 6)$, $(7\ \ 8\ \ 9)$ commute, $\tau^m = (1\ \ 2)^m(3\ \ 4\ \ 5\ \ 6)^m(7\ \ 8\ \ 9)^m$; in order that $\tau^m = e$ we need $(1\ \ 2)^m = e$, $(3\ \ 4\ \ 5\ \ 6)^m = e$, $(7\ \ 8\ \ 9)^m = e$. (Prove!) To have $(7\ \ 8\ \ 9)^m = e$, we must have $3 \mid m$, since $(7\ \ 8\ \ 9)$ is of order 3; to have $(3\ \ 4\ \ 5\ \ 6)^m = e$, we must have $4 \mid m$, because $(3\ \ 4\ \ 5\ \ 6)$ is of order 4, and to have $(1\ \ 2)^m = e$, we must have $2 \mid m$, because $(1\ \ 2)$ is of order 2. This tells us that m must be divisible by 12.

On the other hand,

$$\tau^{12} = (1\ \ 2)^{12}(3\ \ 4\ \ 5\ \ 6)^{12}(7\ \ 8\ \ 9)^{12} = e.$$

So τ is of order 12.

Here, again, the special properties of τ do not enter the picture. What we did for τ works for any permutation. To formulate this properly, recall that the least common multiple of m and n is the smallest positive integer v which is divisible by m and by n. (See Problem 7, Chapter 1, Section 5.) Then we have

Theorem 3.2.4. Let $\sigma \in S_n$ have its cycle decomposition into *disjoint* cycles of length $m_1, m_2, \ldots, m_k$. Then the order of σ is the least common multiple of $m_1, m_2, \ldots, m_k$.

Proof. Let $\sigma = \tau_1\tau_2 \cdots \tau_k$, where the τ_i are disjoint cycles of length m_i. Since the τ_i are disjoint cycles, $\tau_i\tau_j = \tau_j\tau_i$; therefore if M is the least common multiple of $m_1, m_2, \ldots, m_k$, then $\sigma^M = (\tau_1\tau_2 \cdots \tau_k)^M = \tau_1^M\tau_2^M \cdots \tau_k^M = e$ (since $\tau_i^M = e$ because τ_i is of order m_i and $m_i \mid M$). Therefore, the order of σ is *at most* M. On the other hand, if $\sigma^N = e$, then $\tau_1^N\tau_2^N \cdots \tau_k^N = e$. This forces each $\tau_i^N = e$, (prove!) because τ_i are disjoint permutations, so $m_i \mid N$, since τ_i is of order m_i. Thus N is divisible by the least common multiple of $m_1, m_2, \ldots, m_k$, so $M \mid N$. Consequently, we see that σ is of order M as claimed in the theorem. □

Note that the disjointness of the cycles in the theorem is imperative. For instance, (1 2) and (1 3), which are not disjoint, are each of order 2, but their product (1 2)(1 3) = (1 3 2) is of order 3.

Let's consider Theorem 3.2.4 in the context of a card shuffle. Suppose that we shuffle a deck of 13 cards in such a way that the top card is put into the position of the 3rd card, the second in that of the 4th, ..., the ith into the $i + 2$ position, working mod 13. As a permutation, σ, of 1, 2, ..., 13, the shuffle becomes

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\ 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 1 & 2 \end{pmatrix},$$

and σ is merely the 13-cycle (1 3 5 7 9 11 13 2 4 6 8 10 12), so σ is of order 13. How many times must we repeat this shuffle to get the cards back to their original order? The answer is merely the order of σ, that is, 13. So it takes 13 repeats of the shuffle to get the cards back to their original order.

Let's give a twist to the shuffle above. Suppose that we shuffle the cards as follows. First take the top card and put it into the second-to-last place, and then follow it by the shuffle given above. How many repeats are now needed to get the cards back to their original order? The first operation is the shuffle given by the permutation τ = (1 12 11 10 9 8 7 6 5 4 3 2) followed by σ above. So we must compute $\sigma\tau$ and find its order. But

$$\begin{aligned} \sigma\tau &= (1\ 3\ 5\ 7\ 9\ 11\ 13\ 2\ 4\ 6\ 8\ 10\ 12) \\ &\quad \times (1\ 12\ 11\ 10\ 9\ 8\ 7\ 6\ 5\ 4\ 3\ 2) \\ &= (1)(2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ 13), \end{aligned}$$

so is of order 12. So it would take 12 repeats of the shuffle to get back to the original order.

Can you find a shuffle of the 13 cards that would require 42 repeats? Or 20 repeats? What shuffle would require the greatest number of repeats, and what would this number be?

We return to the general discussion. Consider the permutation $(1\ \ 2\ \ 3)$; we see that $(1\ \ 2\ \ 3) = (1\ \ 3)(1\ \ 2)$. We can also see that $(1\ \ 2\ \ 3) = (2\ \ 3)(1\ \ 3)$. So two things are evident. First, we can write $(1\ \ 2\ \ 3)$ as the product of two transpositions, and in at least two distinct ways. Given the k-cycle $(i_1\ \ i_2\ \cdots\ i_k)$, then $(i_1\ \ i_2\ \cdots\ i_k) = (i_1\ \ i_k)(i_1\ \ i_{k-1})\cdots(i_1\ \ i_2)$, so every k-cycle is a product of $k-1$ transpositions (if $k > 1$) and this can be done in several ways, so not in a unique way. Because every permutation is the product of disjoint cycles and every cycle is a product of transpositions we have

Theorem 3.2.5. Every permutation in S_n is the product of transpositions.

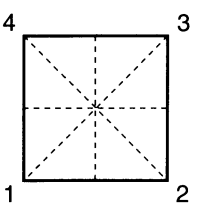
This theorem is really not surprising for it says, after all, nothing more or less than that any permutation can be effected by carrying out a series of interchanges of two objects at a time.

We saw that there is a lack of uniqueness in representing a given permutation as a product of transpositions. But, as we shall see in Section 3, some aspects of this decomposition are indeed unique.

As a final word of this section, we would like to point out the convenience of cycle notation. When we represent elements of a permutation group as products of disjoint cycles, many things become transparent—for example, the order of the permutation is visible at a glance. To illustrate this point, we now give a few examples of certain geometric groups, which are in fact permutation groups that have already appeared in Chapter 2 under different guises.

Examples

1. Informally, a *motion* of a geometric figure is a permutation of its vertices that can be realized by a rigid motion in space. For example, there are eight motions of a square, whose vertices are numbered 1, 2, 3, 4 as below:



$\alpha = (13)$ is the reflection about the axis of symmetry joining vertices 2 and 4 in the original position,

$\beta = (1234)$ is the counterclockwise rotation by 90°,

$\beta^2 = (13)(24)$ is the counterclockwise rotation by 180°,

$\beta^3 = (1432)$ is the counterclockwise rotation by 270°,

$\alpha\beta = (12)(34)$ is the reflection in the vertical axis of symmetry,

$\alpha\beta^2 = (24)$ is the reflection in the other diagonal axis,

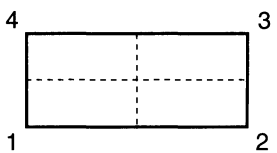
$\alpha\beta^3 = (14)(23)$ is the reflection in the horizontal axis, and, of course

$\alpha^2 = \beta^4 = (1)$ is the "motion" that leaves the vertices unchanged.

We also have the relation $\beta\alpha = \alpha\beta^3$.

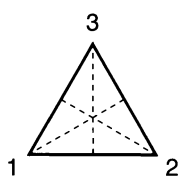
These motions, or *symmetries* of a square, form a subgroup of S_4 which is called the *octic group*, or the *dihedral group* of order 8. This group (or, strictly speaking, a group isomorphic to it) was introduced in Example 9 of Section 2.1 without mention of permutations.

2. There are only four symmetries of a non–square rectangle:



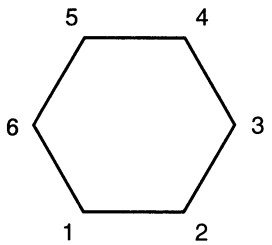
the reflections in the two axes of symmetry, rotation by 180° and the identity. These motions can be identified with permutations (1), (14)(23), (12)(34), (13)(24), and form a subgroup of the group obtained in Example 1. This subgroup is often called *Klein's 4-group.*

3. We leave it to the reader to verify that the group of all motions of an equilateral triangle is the full symmetric group S_3.



4. The motions of a regular hexagon form the dihedral group of order 12, generated by the permutations $\alpha = (15)(24)$, corresponding to a reflection

about one of the axes of symmetry, and $\beta = (123456)$, corresponding to the counterclockwise rotation by 60°.



In general, the dihedral group of order $2n$, which was first introduced in Example 10 of Section 2.1, can be interpreted as the group of symmetries of a regular n-gon (a polygon with n edges of equal length).

PROBLEMS

Easier Problems

1. Show that if σ, τ are two disjoint cycles, then $\sigma\tau = \tau\sigma$.
2. Find the cycle decomposition and order.
   (a) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 1 & 4 & 2 & 7 & 6 & 9 & 8 & 5 \end{pmatrix}$.
   (b) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 6 & 5 & 4 & 3 & 2 & 1 \end{pmatrix}$.
   (c) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 6 & 5 & 3 & 4 & 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 1 & 5 & 6 & 7 & 4 \end{pmatrix}$.
3. Express as the product of disjoint cycles and find the order.
   (a) $(1\ 2\ 3\ 5\ 7)(2\ 4\ 7\ 6)$.
   (b) $(1\ 2)(1\ 3)(1\ 4)$.
   (c) $(1\ 2\ 3\ 4\ 5)(1\ 2\ 3\ 4\ 6)(1\ 2\ 3\ 4\ 7)$.
   (d) $(1\ 2\ 3)(1\ 3\ 2)$.
   (e) $(1\ 2\ 3)(3\ 5\ 7\ 9)(1\ 2\ 3)^{-1}$.
   (f) $(1\ 2\ 3\ 4\ 5)^3$.
4. Give a complete proof of Theorem 3.2.2.
5. Show that a k-cycle has order k.
6. Find a shuffle of a deck of 13 cards that requires 42 repeats to return the cards to their original order.

7. Do Problem 6 for a shuffle requiring 20 repeats.
8. Express the permutations in Problem 3 as the product of transpositions.
9. Given the two transpositions $(1\ \ 2)$ and $(1\ \ 3)$, find a permutation σ such that $\sigma(1\ \ 2)\sigma^{-1} = (1\ \ 3)$.
10. Prove that there is no permutation σ such that $\sigma(1\ \ 2)\sigma^{-1} = (1\ \ 2\ \ 3)$.
11. Prove that there is a permutation σ such that $\sigma(1\ \ 2\ \ 3)\sigma^{-1} = (4\ \ 5\ \ 6)$.
12. Prove that there is no permutation σ such that $\sigma(1\ \ 2\ \ 3)\sigma^{-1} = (1\ \ 2\ \ 4)(5\ \ 6\ \ 7)$.

Middle-Level Problems

13. Prove that $(1\ \ 2)$ cannot be written as the product of disjoint 3-cycles.
14. Prove that for any permutation σ, $\sigma\tau\sigma^{-1}$ is a transposition if τ is a transposition.
15. Show that if τ is a k-cycle, then $\sigma\tau\sigma^{-1}$ is also a k-cycle, for any permutation σ.
16. Let Φ be an automorphism of S_3. Show that there is an element $\sigma \in S_3$ such that $\Phi(\tau) = \sigma^{-1}\tau\sigma$ for every $\tau \in S_3$.
17. Let $(1\ \ 2)$ and $(1\ \ 2\ \ 3\ \ \cdots\ \ n)$ be in S_n. Show that any subgroup of S_n that contains both of these must be all of S_n (so these two permutations generate S_n).
18. If τ_1 and τ_2 are two transpositions, show that $\tau_1\tau_2$ can be expressed as the product of 3-cycles (not necessarily disjoint).
19. Prove that if τ_1, τ_2, and τ_3 are transpositions, then $\tau_1\tau_2\tau_3 \neq e$, the identity element of S_n.
20. If τ_1, τ_2 are distinct transpositions, show that $\tau_1\tau_2$ is of order 2 or 3.
21. If σ, τ are two permutations that disturb no common element and $\sigma\tau = e$, prove that $\sigma = \tau = e$.
22. Find an algorithm for finding $\sigma\tau\sigma^{-1}$ for any permutations σ, τ of S_n.
23. Let σ, τ be two permutations such that they both have decompositions into disjoint cycles of cycles of lengths $m_1, m_2, \ldots, m_k$. (We say that they have similar decompositions into disjoint cycles.) Prove that for some permutation ρ, $\tau = \rho\sigma\rho^{-1}$.
24. Find the conjugacy class in S_n of $(1\ \ 2\ \ \cdots\ \ n)$. What is the order of the centralizer of $(1\ \ 2\ \ \cdots\ \ n)$ in S_n?
25. Do Problem 24 for $\sigma = (1\ \ 2)(3\ \ 4)$.

3. ODD AND EVEN PERMUTATIONS

We noticed in Section 2 that although every permutation is the product of transpositions, this decomposition is not unique. We did comment, however, that certain aspects of this kind of decomposition are unique. We go into this now.

Let's consider the special case of S_3, for here we can see everything explicitly. Let $f(x_1, x_2, x_3) = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$ be an expression in the three variables x_1, x_2, x_3. We let S_3 act on $f(x) = f(x_1, x_2, x_3)$ as follows. If $\sigma \in S_3$, then

$$\sigma^*(f(x)) = (x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)}).$$

We consider what σ^* does to $f(x)$ for a few of the σ's in S_3.

Consider $\sigma = (1 \quad 2)$. Then $\sigma(1) = 2$, $\sigma(2) = 1$, and $\sigma(3) = 3$, so that

$$\begin{aligned}\sigma^*(f(x)) &= (x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)}) \\ &= (x_2 - x_1)(x_2 - x_3)(x_1 - x_3) \\ &= -(x_1 - x_2)(x_1 - x_3)(x_2 - x_3) \\ &= -f(x).\end{aligned}$$

So σ^* coming from $\sigma = (1 \quad 2)$ changes the sign of $f(x)$. Let's look at the action of another element, $\tau = (1 \quad 2 \quad 3)$, of S_3 on $f(x)$. Then

$$\begin{aligned}\tau^*(f(x)) &= (x_{\tau(1)} - x_{\tau(2)})(x_{\tau(1)} - x_{\tau(3)})(x_{\tau(2)} - x_{\tau(3)}) \\ &= (x_2 - x_3)(x_2 - x_1)(x_3 - x_1) \\ &= (x_1 - x_2)(x_1 - x_3)(x_2 - x_3) \\ &= f(x),\end{aligned}$$

so τ^* coming from $\tau = (1 \quad 2 \quad 3)$ leaves $f(x)$ unchanged. What about the other permutations in S_3; how do they affect $f(x)$? Of course, the identity element e induces a map e^* on $f(x)$ which does not change $f(x)$ at all. What does τ^2, for τ above, do to $f(x)$? Since $\tau^* f(x) = f(x)$, we immediately see that

$$\begin{aligned}(\tau^2) * (f(x)) &= (x_{\tau^2(1)} - x_{\tau^2(2)})(x_{\tau^2(1)} - x_{\tau^2(3)})(x_{\tau^2(2)} - x_{\tau^2(3)}) \\ &= f(x). \quad \text{(Prove!)}\end{aligned}$$

Now consider $\sigma\tau = (1 \quad 2)(1 \quad 2 \quad 3) = (2 \quad 3)$; since τ leaves $f(x)$ alone and σ changes the sign of $f(x)$, $\sigma\tau$ must change the sign of $f(x)$. Similarly, $(1 \quad 3)$ changes the sign of $f(x)$. We have accounted for the action of every element of S_3 on $f(x)$.

Suppose that $\rho \in S_3$ is a product $\rho = \tau_1\tau_2 \cdots \tau_k$ of transpositions $\tau_1, \ldots, \tau_k$; then ρ acting on $f(x)$ will change the sign of $f(x)$ k times, since each τ_i changes the sign of $f(x)$. So $\rho^*(f(x)) = (-1)^k f(x)$. If $\rho = \sigma_1\sigma_2 \cdots \sigma_t$, where $\sigma_1, \ldots, \sigma_t$ are transpositions, by the same reasoning, $\rho^*(f(x)) = (-1)^t f(x)$. Therefore, $(-1)^k f(x) = (-1)^t f(x)$, whence $(-1)^t = (-1)^k$. This tells us that t and k have the *same parity*; that is, if t is odd, then k must be odd, and if t is even, then k must be even.

This suggests that although the decomposition of a given permutation σ as a product of transposition is not unique, *the parity of the number of transpositions in such a decomposition of σ might be unique.*

We strive for this goal now, suggesting to readers that they carry out the argument that we do for arbitrary n for the special case $n = 4$.

As we did above, define $f(x) = f(x_1, \ldots, x_n)$ to be

$$f(x) = (x_1 - x_2) \cdots (x_1 - x_n)(x_2 - x_3) \cdots (x_2 - x_n) \cdots (x_{n-1} - x_n)$$

$$= \prod_{i<j} (x_i - x_j),$$

where in this product i takes on all values from 1 to $n - 1$ inclusive, and j all those from 2 to n inclusive. If $\sigma \in S_n$, define σ^* on $f(x)$ by

$$\sigma^*(f(x)) = \prod_{i<j} (x_{\sigma(i)} - x_{\sigma(j)}).$$

If $\sigma, \tau \in S_n$, then

$$(\sigma\tau)^*(f(x)) = \prod_{i<j} (x_{(\sigma\tau)(i)} - x_{(\sigma\tau)(j)}) = \sigma^*\left(\prod_{i<j} (x_{\tau(i)} - x_{\tau(j)})\right)$$

$$= \sigma^*\left(\tau^*\left(\prod_{i<j} (x_i - x_j)\right)\right) = \sigma^*(\tau^*(f(x))) = (\sigma^*\tau^*)(f(x))$$

So $(\sigma\tau)^* = \sigma^*\tau^*$ when applied to $f(x)$.

What does a transposition τ do to $f(x)$? We claim that $\tau^*(f(x)) = -f(x)$. To prove this, assuming that $\tau = (i \quad j)$ where $i < j$, we count up the number of $(x_u - x_v)$, with $u < v$, which get transformed into an $(x_a - x_b)$ with $a > b$. This happens for $(x_u - x_j)$ if $i < u < j$, for $(x_i - x_v)$ if $i < v < j$, and finally, for $(x_i - x_j)$. Each of these leads to a change of sign on $f(x)$ and since there are $2(j - i - 1) + 1$ such, that is, an odd number of them, we get an odd number of changes of sign on $f(x)$ when acted on by τ^*. Thus $\tau^*(f(x)) = -f(x)$. Therefore, our claim that $\tau^*(f(x)) = -f(x)$ for every transposition τ is substantiated.

If σ is any permutation in S_n and $\sigma = \tau_1\tau_2 \cdots \tau_k$, where $\tau_1, \tau_2, \ldots, \tau_k$ are transpositions, then $\sigma^* = (\tau_1\tau_2 \cdots \tau_k)^* = \tau_1^*\tau_2^* \cdots \tau_k^*$ as acting on $f(x)$, and since each $\tau_i^*(f(x)) = -f(x)$, we see that $\sigma^*(f(x)) = (-1)^k f(x)$. Similarly, if $\sigma = \zeta_1\zeta_2 \cdots \zeta_t$, where $\zeta_1, \zeta_2, \ldots, \zeta_t$ are transpositions, then $\sigma^*(f(x)) = (-1)^t f(x)$. Comparing these two evaluations of $\sigma^*(f(x))$, we conclude that $(-1)^k = (-1)^t$. So these two decompositions of σ as the product of transpositions are of the same parity. *Thus any permutation is either the product of an odd number of transpositions or the product of an even number of transpositions, and no product of an even number of transpositions can equal a product of an odd number of transpositions.*

This suggests the following

Definition. The permutation $\sigma \in S_n$ is an *odd permutation* if σ is the product of an odd number of transpositions, and is an *even permutation* if σ is the product of an even number of transpositions.

What we have proved above is

Theorem 3.3.1. A permutation in S_n is either an odd or an even permutation, *but cannot be both.*

With Theorem 3.3.1 behind us we can deduce a number of its consequences.

Let A_n be the set of all even permutations; if $\sigma, \tau \in A_n$, then we immediately have that $\sigma\tau \in A_n$. Since A_n is thus a finite closed subset of the (finite) group S_n, A_n is a subgroup of S_n, by Lemma 2.3.2. A_n is called the *alternating group of degree n.*

We can show that A_n is a subgroup of S_n in another way. We already saw that A_n is closed under the product of S_n, so to know that A_n is a subgroup of S_n we merely need show that $\sigma \in S_n$ implies that $\sigma^{-1} \in S_n$. For any permutation σ we claim that σ and σ^{-1} are of the same parity. Why? Well, if $\sigma = \tau_1\tau_2 \cdots \tau_k$, where the τ_i are transpositions, then

$$\sigma^{-1} = (\tau_1\tau_2 \cdots \tau_k)^{-1} = \tau_k^{-1}\tau_{k-1}^{-1} \cdots \tau_2^{-1}\tau_1^{-1} = \tau_k\tau_{k-1} \cdots \tau_2\tau_1,$$

since $\tau_i^{-1} = \tau_i$. Therefore, we see that the parity of σ and σ^{-1} is $(-1)^k$, so they are of equal parity. This certainly shows that $\sigma \in A_n$ forces $\sigma^{-1} \in A_n$, whence A_n is a subgroup of S_n.

But it shows a little more, namely that A_n is a *normal subgroup* of S_n. For suppose that $\sigma \in A_n$ and $\rho \in S_n$. What is the parity of $\rho^{-1}\sigma\rho$? By the

above, ρ and ρ^{-1} are of the same parity and σ is an even permutation so $\rho^{-1}\sigma\rho$ is an even permutation, hence is in A_n. Thus A_n is a normal subgroup of S_n.

We summarize what we have done in

Theorem 3.3.2. A_n, the alternating group of degree n, is a normal subgroup of S_n.

We look at this in yet another way. From the very definitions involved we have the following simple rules for the product of permutations:

1. The product of two even permutations is even.
2. The product of two odd permutations is even.
3. The product of an even permutation by an odd one (or of an odd one by an even one) is odd.

If σ is an even permutation, let $\theta(\sigma) = 1$, and if σ is an odd permutation, let $\theta(\sigma) = -1$. The foregoing rules about products translate into $\theta(\sigma\tau) = \theta(\sigma)\theta(\tau)$, so θ is a *homomorphism* of S_n onto the group $E = \{1, -1\}$ of order 2 under multiplication. What is the kernel, N, of θ? By the very definition of A_n we see that $N = A_n$. So by the First Homomorphism Theorem, $E \simeq S_n/A_n$. Thus $2 = |E| = |S_n/A_n| = |S_n|/|A_n|$, if $n > 1$. This gives us that $|A_n| = \frac{1}{2}|S_n| = \frac{1}{2}n!$.

Therefore,

Theorem 3.3.3. For $n > 1$, A_n is a normal subgroup of S_n of order $\frac{1}{2}n!$.

Corollary. For $n > 1$, S_n contains $\frac{1}{2}n!$ even permutations and $\frac{1}{2}n!$ odd permutations.

A final few words about the proof of Theorem 3.3.1 before we close this section. Many different proofs of Theorem 3.3.1 are known. Quite frankly, we do not particularly like any of them. Some involve what might be called a "collection process," where one tries to show that e cannot be written as the product of an odd number of transpositions by assuming that it is such a shortest product, and by the appropriate finagling with this product, shortening it to get a contradiction. Other proofs use other devices. The proof we gave exploits the gimmick of the function $f(x)$, which, in some sense, is extraneous to the whole affair. However, the proof given is probably the most transparent of them all, which is why we used it.

Finally, the group A_n, for $n \geq 5$, is an extremely interesting group. We

shall show in Chapter 6 that the only normal subgroups of A_n, for $n \geq 5$, are (e) and A_n itself. A group with this property is called a *simple* group (not to be confused with an *easy* group). The abelian finite simple groups are merely the groups of prime order. The A_n for $n \geq 5$ provide us with an infinite family of *nonabelian* finite simple groups. There are other infinite families of finite simple groups. In the last 20 years or so the heroic efforts of algebraists have determined all finite simple groups. The determination of these simple groups runs about 10,000 printed pages. Interestingly enough, any nonabelian finite simple group must have *even* order.

PROBLEMS

Easier Problems

1. Find the parity of each permutation.
   (a) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 4 & 5 & 1 & 3 & 7 & 8 & 9 & 6 \end{pmatrix}$.
   (b) $(1\ 2\ 3\ 4\ 5\ 6)(7\ 8\ 9)$.
   (c) $(1\ 2\ 3\ 4\ 5\ 6)(1\ 2\ 3\ 4\ 5\ 7)$.
   (d) $(1\ 2)(1\ 2\ 3)(4\ 5)(5\ 6\ 8)(1\ 7\ 9)$.
2. If σ is a k-cycle, show that σ is an odd permutation if k is even, and is an even permutation if k is odd.
3. Prove that σ and $\tau^{-1}\sigma\tau$, for any $\sigma, \tau \in S_n$, are of the same parity.
4. If $m < n$, we can consider $S_m \subset S_n$ by viewing $\sigma \in S_n$ as acting on $1, 2, \ldots, m, \ldots, n$ as it did on $1, 2, \ldots, m$ and σ leaves $j > m$ fixed. Prove that the parity of a permutation in S_m, when viewed this way as an element of S_n, does not change.
5. Suppose you are told that the permutation
   $$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 1 & 2 & & & 7 & 8 & 9 & 6 \end{pmatrix}$$
   in S_9, where the images of 5 and 4 have been lost, is an even permutation. What must the images of 5 and 4 be?

Middle-Level Problems

6. If $n \geq 3$, show that every element in A_n is a product of 3-cycles.
7. Show that every element in A_n is a product of n-cycles.
8. Find a normal subgroup in A_4 of order 4.

Harder Problems (In fact, very hard)

9. If $n \geq 5$ and $(e) \neq N \subset A_n$ is a normal subgroup of A_n, show that N must contain a 3-cycle.

10. Using the result of Problem 9, show that if $n \geq 5$, the only normal subgroups of A_n are (e) and A_n itself. (Thus the groups A_n for $n \geq 5$ give us an infinite family of nonabelian finite simple groups.)

4

RING THEORY

Ziyad Asaad

1. DEFINITIONS AND EXAMPLES

So far in our study of abstract algebra, we have been introduced to one kind of abstract system, which plays a central role in the algebra of today. That was the notion of a group. Because a group is an algebraic system with only one operation, and because a group need not satisfy the rule $ab = ba$, it ran somewhat counter to our prior experience in algebra. We were used to systems where you could both add and multiply elements and where the elements did satisfy the commutative law of multiplication $ab = ba$. Furthermore, these systems of our acquaintance usually came from sets of numbers—integers, rational, real, and for some, complex.

The next algebraic object we shall consider is a *ring*. In many ways this system will be more reminiscent of what we had previously known than were groups. For one thing rings will be endowed with addition and multiplication, and these will be subjected to many of the familiar rules we all know from arithmetic. On the other hand, rings need not come from our usual number systems, and, in fact, usually have little to do with these familiar ones. Although many of the formal rules of arithmetic hold, many strange—or what may seem as strange—phenomena do take place. As we proceed and see examples of rings, we shall see some of these things occur.

With this preamble over we are ready to begin. Naturally enough, the first thing we should do is to define that which we'll be talking about.

Definition. A nonempty set R is said to be a *ring* if in R there are two operations $+$ and $\cdot$ such that:

(a) $a, b \in R$ implies that $a + b \in R$.
(b) $a + b = b + a$ for $a, b \in R$.
(c) $(a + b) + c = a + (b + c)$ for $a, b, c \in R$.
(d) There exists an element $0 \in R$ such that $a + 0 = a$ for every $a \in R$.
(e) Given $a \in R$, there exists a $b \in R$ such that $a + b = 0$. (We shall write b as $-a$.)

Note that so far all we have said is that R is an abelian group under $+$. We now spell out the rules for the multiplication in R.

(f) $a, b \in R$ implies that $a \cdot b \in R$.
(g) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for $a, b, c \in R$.

This is all that we insist on as far as the multiplication by itself is concerned. But the $+$ and $\cdot$ are not allowed to live in solitary splendor. We interweave them by the two *distributive laws*

(h) $a \cdot (b + c) = a \cdot b + a \cdot c$ and
$(b + c) \cdot a = b \cdot a + c \cdot a$, for $a, b, c \in R$.

These axioms for a ring look familiar. They should be, for the concept of ring was introduced as a generalization of what happens in the integers. Because of Axiom (g), the associative law of multiplication, the rings we defined are usually called *associative rings*. Nonassociative rings do exist, and some of these play an important role in mathematics. But they shall not be our concern here. So whenever we use the word "ring" we shall always mean "associative ring."

Although Axioms (a) to (h) are familiar, there are certain things they do not say. We look at some of the familiar rules that are *not insisted* upon for a general ring.

First, we do not postulate the existence of an element $1 \in R$ such that $a \cdot 1 = 1 \cdot a = a$ for every $a \in R$. Many of the examples we shall encounter will have such an element, and in that case we say that *R is a ring with unit.* In all fairness we should point out that many algebraists do demand that a ring have a unit element. We do insist that $1 \neq 0$; that is, the ring consisting of 0 alone is not a ring with unit.

Second, in our previous experience with things of this sort, whenever $a \cdot b = 0$ we concluded that $a = 0$ or $b = 0$. This need not be true, in general, in a ring. When it does hold, the ring is kind of nice and is given a special name; it is called a *domain*.

Third, nothing is said in the axioms for a ring that will imply the *commutative law* of multiplication $a \cdot b = b \cdot a$. There are noncommutative rings where this law does not hold; we shall see some soon. Our main concern in this chapter will be with *commutative rings*, but for many of the early results the commutativity of the ring studied will not be assumed.

As we mentioned above, some things make certain rings nicer than others, and so become worthy of having a special name. We quickly give a list of definitions for some of these nicer rings.

Definition. A commutative ring R is an *integral domain* if $a \cdot b = 0$ in R implies that $a = 0$ or $b = 0$.

It should be pointed out that some algebra books insist that an integral domain contain a unit element. In reading another book, the reader should check if this is the case there. The integers, $\mathbb{Z}$, give us an obvious example of an integral domain. We shall see other, somewhat less obvious ones.

Definition. A ring R with unit is said to be a *division ring* if for every $a \neq 0$ in R there is an element $b \in R$ (usually written as a^{-1}) such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

The reason for calling such a ring a division ring is quite clear, for we can divide (at least keeping left and right sides in mind). Although noncommutative division rings exist with fair frequency and do play an important role in noncommutative algebra, they are fairly complicated and we shall give only one example of these. This division ring is the great classic one introduced by Hamilton in 1843 and is known as the ring of *quaternions*. (See Example 13 below.)

Finally, we come to perhaps the nicest example of a class of rings, the *field*.

Definition. A ring R is said to be a *field* if R is a *commutative division ring*.

In other words, a field is a commutative ring in which we can divide freely by nonzero elements. Otherwise put, R is a field if the nonzero elements of R form an abelian group under $\cdot$, the product in R.

For fields we do have some ready examples: the rational numbers, the real numbers, the complex numbers. But we shall see many more, perhaps less familiar, examples. Chapter 5 will be devoted to the study of fields.

We spend the rest of the time in this section looking at some examples of rings. *We shall drop the · for the product and shall write a · b simply as ab.*

Examples

1. It is obvious which ring we should pick as our first example, namely $\mathbb{Z}$, the ring of integers under the usual addition and multiplication of integers. Naturally enough, $\mathbb{Z}$ is an example of an integral domain.

2. The second example is equally obvious as a choice. Let $\mathbb{Q}$ be the set of all rational numbers. As we all know, $\mathbb{Q}$ satisfies all the rules needed for a field, so $\mathbb{Q}$ is a field.

3. The real numbers, $\mathbb{R}$, also give us an example of a field.

4. The complex numbers, $\mathbb{C}$, form a field.

Note that $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$; we describe this by saying that $\mathbb{Q}$ is a *subfield* of $\mathbb{R}$ (and of $\mathbb{C}$) and $\mathbb{R}$ is a subfield of $\mathbb{C}$.

5. Let $R = \mathbb{Z}_6$, the integers mod 6, with the addition and the multiplication defined by $[a] + [b] = [a + b]$ and $[a][b] = [ab]$.

Note that [0] is the 0 required by our axioms for a ring, and [1] is the unit element of R. Note, however, that $\mathbb{Z}_6$ is *not an integral domain*, for $[2][3] = [6] = [0]$, yet $[2] \neq [0]$ and $[3] \neq [0]$. R is a commutative ring with unit.

This example suggests the

Definition. An element $a \neq 0$ in a ring R is a *zero-divisor* in R if $ab = 0$ for some $b \neq 0$ in R.

We should really call what we defined a left zero-divisor; however, since we shall mainly talk about commutative rings, we shall not need any left-right distinction for zero-divisors.

Note that both [2] and [3] in $\mathbb{Z}_6$ are zero-divisors. An integral domain is, of course, a commutative ring without zero-divisors.

6. Let $R = \mathbb{Z}_5$, the ring of integers mod 5. R is, of course, a commutative ring with unit. But it is more; in fact, it is a field. Its nonzero elements are [1], [2], [3], [4] and we note that $[2][3] = [6] = [1]$, and [1] and [4] are their own inverses. So every nonzero element in $\mathbb{Z}_5$ has an inverse in $\mathbb{Z}_5$.

We generalize this to any prime p.

7. Let $\mathbb{Z}_p$ be the integers mod p, where p is a prime. Again $\mathbb{Z}_p$ is clearly a commutative ring with 1. We claim that $\mathbb{Z}_p$ is a field. To see this, note that if $[a] \neq [0]$, then $p \nmid a$. Therefore, by Fermat's Theorem (Corollary to Theorem 2.4.8), $a^{p-1} \equiv 1(p)$. For the classes $[\cdot]$ this says that $[a^{p-1}] = [1]$. But $[a^{p-1}] = [a]^{p-1}$, so $[a]^{p-1} = [1]$; therefore, $[a]^{p-2}$ is the required inverse for $[a]$ in $\mathbb{Z}_p$, hence $\mathbb{Z}_p$ is a field.

Because $\mathbb{Z}_p$ has only a finite number of elements, it is called a *finite field.* Later we shall construct finite fields different from the $\mathbb{Z}_p$'s.

8. Let $\mathbb{Q}$ be the rational numbers; if $a \in \mathbb{Q}$, we can write $a = m/n$, where m and n are relatively prime integers. Call this the *reduced form* for a. Let R be the set of all $a \in \mathbb{Q}$ in whose reduced form the denominator is odd. Under the usual addition and multiplication in $\mathbb{Q}$ the set R forms a ring. It is an integral domain with unit but is not a field, for $\frac{1}{2}$, the needed inverse of 2, is not in R. Exactly which elements in R do have their inverses in R?

9. Let R be the set of all $a \in \mathbb{Q}$ in whose reduced form the denominator is not divisible by a fixed prime p. As in (8), R is a ring under the usual addition and multiplication in $\mathbb{Q}$, is an integral domain but is not a field. What elements of R have their inverses in R?

Both Examples 8 and 9 are subrings of $\mathbb{Q}$ in the following sense.

Definition. If R is a ring, then a *subring* of R is a subset S of R which is a ring if the operations ab and $a + b$ are just the operations of R applied to the elements $a, b \in S$.

For S to be a subring, it is necessary and sufficient that S be nonempty and that $ab, a \pm b \in S$ for all $a, b \in S$. (Prove!)

We give one further commutative example. This one comes from the calculus.

10. Let R be the set of all real-valued continuous functions on the closed unit interval $[0, 1]$. For $f, g \in R$ and $x \in [0, 1]$ define $(f + g)(x) = f(x) + g(x)$, and $(f \cdot g)(x) = f(x)g(x)$. From the results in the calculus, $f + g$ and $f \cdot g$ are again continuous functions on $[0, 1]$. With these operations R is a commutative ring. It is *not* an integral domain. For instance, if $f(x) = -x + \frac{1}{2}$ for $0 \leq x \leq \frac{1}{2}$ and $f(x) = 0$ for $\frac{1}{2} < x \leq 1$, and if $g(x) = 0$ for $0 \leq x \leq \frac{1}{2}$ and $g(x) = 2x - 1$ for $\frac{1}{2} < x \leq 1$, then $f, g \in R$ and, as is easy to verify, $f \cdot g = 0$. It does have a unit element, namely the function e defined by $e(x) = 1$ for all $x \in [0, 1]$. What elements of R have their inverses in R?

We should now like to see some noncommutative examples. These are not so easy to come by, although noncommutative rings exist in abundance, because we are not assuming any knowledge of linear algebra on the reader's part. The easiest and most natural first source of such examples is the set of matrices over a field. So, in our first noncommutative example, we shall really create the 2×2 matrices with real entries.

11. Let F be the field of real numbers and let R be the set of all formal square arrays

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

where a, b, c, d are any real numbers. For such square arrays we define addition in a natural way by defining

$$\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} + \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = \begin{pmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{pmatrix}.$$

It is easy to see that R forms an abelian group under this $+$ with $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ acting as the zero element and $\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$ the negative of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. To make of R a ring, we need a multiplication. We define one in what may seem a highly unnatural way via

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} r & s \\ t & u \end{pmatrix} = \begin{pmatrix} ar + bt & as + bu \\ cr + dt & cs + du \end{pmatrix}.$$

It may be a little laborious, but one can check that with these operations R is a noncommutative ring with $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ acting as its multiplicative unit element. Note that

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

while

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

so

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ are zero-divisors; in fact,

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{so} \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

is a nonzero element whose square is the 0 element of R. This R is known as the *ring of all* 2×2 *matrices over* F, *the real field.*

For those unfamiliar with these matrices, and who see no sense in the product defined for them, let's look at how we do compute the product. To get the top left entry in the product AB, we "multiply" the first row of A by the first column of B, where $A, B \in R$. For the top right entry, it is the first row of A versus the second column of B. The bottom left entry comes from the second row of A versus the first column of B, and finally, the bottom right entry is the second column of A versus the second column of B.

We illustrate with an example: Let

$$A = \begin{pmatrix} 1 & \frac{1}{2} \\ -3 & 2 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \frac{1}{3} & \frac{2}{5} \\ \pi & -\pi \end{pmatrix}.$$

Then the first row of A is $1, \frac{1}{2}$ and the first column of B is $\frac{1}{3}, \pi$; we "multiply" these via $1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \pi = \pi/2 + \frac{1}{3}$, and so on. So we see that

$$AB = \begin{pmatrix} \frac{1}{3} + \pi/2 & \frac{2}{5} - \pi/2 \\ -1 + 2\pi & -\frac{6}{5} - 2\pi \end{pmatrix}.$$

In the problems we shall have many matrix multiplications, so that the reader can acquire some familiarity with this strange but important example.

12. Let R be any ring and let

$$S = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| \, a, b, c, d \in R \right\}$$

with $+$ and $\cdot$ as defined in Example 11. One can verify that S is a ring, also, under these operations. It is called the *ring of* 2×2 *matrices over* R.

Our final example is one of the great classical examples, the *real quaternions*, introduced by Hamilton (as a noncommutative parallel to the complex numbers).

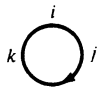
13. *The quaternions.* Let F be the field of real numbers and consider the set of all formal symbols $\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, where $\alpha_0, \alpha_1, \alpha_2, \alpha_3 \in F$. Equality and addition of these symbols are easy, via the obvious route

$$\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 k$$

if and only if $\alpha_0 = \beta_0$, $\alpha_1 = \beta_1$, $\alpha_2 = \beta_2$ and $\alpha_3 = \beta_3$, and

$$(\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k) + (\beta_0 + \beta_1 i + \beta_2 j + \beta_3 k)$$
$$= (\alpha_0 + \beta_0) + (\alpha_1 + \beta_1) i + (\alpha_2 + \beta_2) j + (\alpha_3 + \beta_3) k.$$

We now come to the tricky part, the multiplication. When Hamilton discovered it on October 6, 1843, he cut the basic rules of this product out with his penknife on Brougham Bridge in Dublin. The product is based on $i^2 = j^2 = k^2 = -1$, $ij = k$, $jk = i$, $ki = j$ and $ji = -k$, $kj = -i$, $ik = -j$. If we go around the circle clockwise



the product of any two successive ones is the next one, and going around counterclockwise we get the negatives.

We can write out the product now of any two quaternions, according to the rules above, *declaring by definition* that

$$(\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k)(\beta_0 + \beta_1 i + \beta_2 j + \beta_3 k)$$
$$= \gamma_0 + \gamma_1 i + \gamma_2 j + \gamma_3 k,$$

where

$$\begin{aligned} \gamma_0 &= \alpha_0\beta_0 - \alpha_1\beta_1 - \alpha_2\beta_2 - \alpha_3\beta_3 \\ \gamma_1 &= \alpha_0\beta_1 + \alpha_1\beta_0 + \alpha_2\beta_3 - \alpha_3\beta_2 \\ \gamma_2 &= \alpha_0\beta_2 - \alpha_1\beta_3 + \alpha_2\beta_0 + \alpha_3\beta_1 \\ \gamma_3 &= \alpha_0\beta_3 + \alpha_1\beta_2 - \alpha_2\beta_1 + \alpha_3\beta_0 \end{aligned} \tag{I}$$

It looks horrendous, doesn't it? But it's not as bad as all that. We are multiplying out formally using the distributive laws and using the product rules for the i, j, k above.

If some α_i is 0 in $x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, we shall omit it in expressing x; thus $0 + 0i + 0j + 0k$ will be written simply as 0, $1 + 0i + 0j + 0k$ as 1, $0 + 3i + 4j + 0k$ as $3i + 4j$, and so on.

A calculation reveals that

$$(\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k)(\alpha_0 - \alpha_1 i - \alpha_2 j - \alpha_3 k) \tag{II}$$
$$= \alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2.$$

This has a very important consequence; for suppose that $x = \alpha_0 + a_1 i + \alpha_2 j + \alpha_3 k \neq 0$ (so some $\alpha_i \neq 0$). Then, since the α's are real, $\beta = \alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 \neq 0$. Then from (II) we easily get

$$(\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k)\left(\frac{\alpha_0}{\beta} - \frac{\alpha_1}{\beta} i - \frac{\alpha_2}{\beta} j - \frac{\alpha_3}{\beta} k\right) = 1.$$

So, if $x \neq 0$, then x has an inverse in the quaternions. *Thus the quaternions form a noncommutative division ring.*

Although, as we mentioned earlier, there is no lack of noncommutative division rings, the quaternions above (or some piece of them) are often the only noncommutative division rings that even many professional mathematicians have ever seen.

We shall have many problems—some easy and some quite a bit harder—about the two examples: the 2×2 matrices and the quaternions. This way the reader will be able to acquire some skill with playing with noncommutative rings.

One final comment in this section: If $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ are as in (I), then

$$\begin{aligned}&(\alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2)(\beta_0^2 + \beta_1^2 + \beta_2^2 + \beta_3^2) \\ &= \gamma_0^2 + \gamma_1^2 + \gamma_2^2 + \gamma_3^2.\end{aligned} \tag{III}$$

This is known as *Lagrange's Identity*; it expresses the product of two sums of four squares again as a sum of four squares. Its verification will be one of the exercises.

PROBLEMS

Easier Problems

*1. Find all the elements in $\mathbb{Z}_{24}$ that are *invertible* (i.e., have a multiplicative inverse) in $\mathbb{Z}_{24}$.

2. Show that any field is an integral domain.

3. Show that $\mathbb{Z}_n$ is a field if and only if n is a prime.

4. Verify that Example 8 is a ring. Find all its invertible elements.

5. Do Problem 4 for Example 9.

6. In Example 11, the 2×2 matrices over the reals, check the associative law of multiplication.

7. Work out the following:

(a) $\begin{pmatrix} 1 & 2 \\ 4 & -7 \end{pmatrix}\begin{pmatrix} \frac{1}{5} & \frac{2}{3} \\ 0 & 1 \end{pmatrix}$.

(b) $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}^2$.

(c) $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{pmatrix}^3$.

(d) $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

8. Find all matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

9. Find all 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ that commute with all 2×2 matrices.

10. Let R be any ring with unit, S the ring of 2×2 matrices over R. (See Example 12.)

(a) Check the associative law of multiplication in S. (**Remember:** R need not be commutative.)

(b) Show that $\left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \middle| \, a, b, c, \in R \right\}$ is a subring of S.

(c) Show that $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ has an inverse in S if and only if a and c have inverses in R. In that case write down $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}^{-1}$ explicitly.

11. Let $F: \mathbb{C} \to \mathbb{C}$ be defined by $F(a + bi) = a - bi$. Show that:

(a) $F(xy) = F(x)F(y)$ for $x, y \in \mathbb{C}$.

(b) $F(x\bar{x}) = |x|^2$.

(c) Using Parts (a) and (b), show that $(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2$.

[**Note:** $F(x)$ is merely $\bar{x}$.]

12. Verify the identity in Part (c) of Problem 11 directly.

13. Find the following products of quaternions.

(a) $(i + j)(i - j)$.

(b) $(1 - i + 2j - 2k)(1 + 2i - 4j + 6k)$.

(c) $(2i - 3j + 4k)^2$.

(d) $i(\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k) - (\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k)i$.

14. Show that the only quaternions commuting with i are of the form $\alpha + \beta i$.

15. Find the quaternions that commute with both i and j.

16. Verify that

$$(\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k)(\alpha_0 - \alpha_1 i - \alpha_2 j - \alpha_3 k) = \alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2.$$

17. Verify Lagrange's Identity by a direct calculation.

Middle-Level Problems

18. In the quaternions, define

$$|\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k| = \sqrt{\alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2}.$$

Show that $|xy| = |x|\,|y|$ for any two quaternions x and y.

19. Show that there is an *infinite* number of solutions to $x^2 = -1$ in the quaternions.

20. In the quaternions, consider the following set G having eight elements: $G = \{\pm 1, \pm i, \pm j, \pm k\}$.

(a) Prove that G is a group (under multiplication).

(b) List all subgroups of G.

(c) What is the center of G?

(d) Show that G is a *nonabelian group all of whose subgroups are normal.*

21. Show that a division ring is a domain.

22. Give an example, in the quaternions, of a noncommutative domain that is not a division ring.

23. *Define* the map * in the quaternions by

$$(\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k)^* = (\alpha_0 - \alpha_1 i - \alpha_2 j - \alpha_3 k).$$

Show that:

(a) $x^{**} = (x^*)^* = x$.

(b) $(x + y)^* = x^* + y^*$.

(c) $xx^* = x^*x$ is real and nonnegative.

(d) $(xy)^* = y^*x^*$.

[Note the reversal of order in Part (d).]

24. Using *, define $|x| = \sqrt{xx^*}$. Show that $|xy| = |x|\,|y|$ for any two quaternions x and y, by using Parts (c) and (d) of Problem 23.

25. Use the result of Problem 24 to prove Lagrange's Identity.

In Problems 26 to 30, let R be the 2×2 matrices over the reals.

26. If $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in R$, show that $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible in R if and only if $ad - bc \neq 0$. In that case find $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}$.

27. Define $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$. For $x, y \in R$ show that $\det(xy) = (\det x)(\det y)$.

28. Show that $\{x \in R \mid \det x \neq 0\}$ forms a group, G, under matrix multiplication and that $N = \{x \in R \mid \det x = 1\}$ is a normal subgroup of G.

29. If $x \in R$ is a zero-divisor, show that $\det x = 0$, and, conversely, if $x \neq 0$ is such that $\det x = 0$, then x is a zero-divisor in R.

30. In R, show that $\left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \middle| \, a, b \text{ real} \right\}$ is a field.

Harder Problems

31. Let R be the ring of all 2×2 matrices over $\mathbb{Z}_p$, p a prime. Show that if $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \neq 0$, then $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible in R.

32. Let R be as in Problem 31. Show that for $x, y \in R$, $\det(xy) = \det(x)\det(y)$.

33. Let G be the set of elements x in the ring R of Problem 31 such that $\det(x) \neq 0$.

(a) Prove that G is a group.

(b) Find the order of G. (*Quite hard*)

(c) Find the center of G.

(d) Find a p-Sylow subgroup of G.

34. Let T be the group of matrices A with entries in the field $\mathbb{Z}_2$ such that $\det A$ is not equal to 0. Prove that T is isomorphic to S_3, the symmetric group of degree 3.

35. For R as in Example 10, show that $S = \{f \in R \mid f \text{ is differentiable on } (0, 1)\}$ is a subring of R which is not an integral domain.

If F is a field, let $H(F)$ be the ring of quaternions over F, that is, the set of all $\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, where $\alpha_0, \alpha_1, \alpha_2, \alpha_3 \in F$ and where equality, addition, and multiplication are defined as for the real quaternions.

36. If $F = \mathbb{C}$, the complex numbers, show that $H(\mathbb{C})$ is *not* a division ring.

37. In $H(\mathbb{C})$, find an element $x \neq 0$ such that $x^2 = 0$.

38. Show that $H(F)$ is a division ring if and only if $\alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 0$ for $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ in F forces $\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 0$.

39. If $\mathbb{Q}$ is the field of rational numbers, show that $H(\mathbb{Q})$ is a division ring.

40. Prove that a finite domain is a division ring.

41. Use Problem 40 to show that $\mathbb{Z}_p$ is a field if p is a prime.

2. SOME SIMPLE RESULTS

Now that we have seen some examples of rings and have had some experience playing around with them, it would seem wise to develop some computational rules. These will allow us to avoid annoying trivialities that could beset a calculation we might be making.

The results we shall prove in this section are not very surprising, not too interesting, and certainly not at all exciting. Neither was learning the alphabet, but it was something we had to do before going on to bigger and better things. The same holds for the results we are about to prove.

Since a ring R is at least an abelian group under $+$, there are certain things we know from our group theory background, for instance, $-(-a) = a$, $-(a + b) = (-a) + (-b)$; if $a + b = a + c$, then $b = c$, and so on.

We begin with

Lemma 4.2.1. Let R be any ring and let $a, b \in R$. Then

(a) $a0 = 0a = 0$.

(b) $a(-b) = (-a)b = -(ab)$.

(c) $(-a)(-b) = ab$.

(d) If $1 \in R$, then $(-1)a = -a$.

Proof. We do these in turn.

(a) Since $0 = 0 + 0$, $a0 = a(0 + 0) = a0 + a0$, hence $a0 = 0$. We have used the left distributive law in this proof. The right distributive law gives $0a = 0$.

(b) $ab + a(-b) = a(b + (-b)) = a0 = 0$ from Part (a). Therefore, $a(-b) = -(ab)$. Similarly, $(-a)b = -(ab)$.

(c) By Part (b), $(-a)(-b) = -((-a)b) = -(-(ab)) = ab$, since we are in an abelian group.

(d) If $1 \in R$, then $(-1)a + a = (-1)a + (1)a = (-1 + 1)a = 0a = 0$. So $(-1)a = -a$ by the definition of $-a$. □

Another computational result.

Lemma 4.2.2. In any ring R, $(a + b)^2 = a^2 + b^2 + ab + ba$ for $a, b \in R$.

Proof. This is clearly the analog of $(\alpha + \beta)^2 = \alpha^2 + 2\alpha\beta + \beta^2$ in the integers, say, but keeping in mind that R may be noncommutative. So, to it. By the right distributive law $(a + b)^2 = (a + b)(a + b) = (a + b)a + (a + b)b = a^2 + ba + ab + b^2$, exactly what was claimed. □

Can you see the noncommutative version of the binomial theorem? Try it for $(a + b)^3$.

One curiosity follows from the two distributive laws when R has a unit element. The commutative law of addition follows from the rest.

Lemma 4.2.3. If R is a system with 1 satisfying all the axioms of a ring, except possibly $a + b = b + a$ for $a, b \in R$, then R is a ring.

Proof. We must show that $a + b = b + a$ for $a, b \in R$. By the right distributive law $(a + b)(1 + 1) = (a + b)1 + (a + b)1 = a + b + a + b$. On the other hand, by the left distributive law $(a + b)(1 + 1) = a(1 + 1) + b(1 + 1) = a + a + b + b$. But then $a + b + a + b = a + a + b + b$; since we are in a group under $+$, we can cancel a on the left and b on the right to obtain $b + a = a + b$, as required. R is therefore a ring. □

We close this brief section with a result that is a little nicer. We say that a ring R is a *Boolean ring* [after the English mathematician George Boole (1815–1864)] if $x^2 = x$ for every $x \in R$.

We prove a nice result on Boolean rings.

Lemma 4.2.4. A Boolean ring is commutative.

Proof. Let $x, y \in R$, a Boolean ring. Thus $x^2 = x$, $y^2 = y$, $(x + y)^2 = x + y$. But $(x + y)^2 = x^2 + xy + yx + y^2 = x + xy + yx + y$, by Lemma 4.2.2, so we have $(x + y) = (x + y)^2 = x + xy + yx + y$, from which we have $xy + yx = 0$. Thus $0 = x(xy + yx) = x^2y + xyx = xy + xyx$, while $0 = (xy + yx)x = xyx + yx^2 = xyx + yx$. This gives us $xy + xyx = xyx + yx$, and so $xy = yx$. Therefore, R is commutative. □

PROBLEMS

1. Let R be a ring: since R is an abelian group under $+$, na has meaning for us for $n \in \mathbb{Z}$, $a \in R$. Show that $(na)(mb) = (nm)(ab)$ if n, m are integers and $a, b \in R$.
2. If R is an integral domain and $ab = ac$ for $a \neq 0$, $b, c \in R$, show that $b = c$.
3. If R is a finite integral domain, show that R is a field.
4. If R is a ring and $e \in R$ is such that $e^2 = e$, show that $(xe - exe)^2 = (ex - exe)^2 = 0$ for every $x \in R$.
5. Let R be a ring in which $x^3 = x$ for every $x \in R$. Prove that R is commutative.
6. If $a^2 = 0$ in R, show that $ax + xa$ commutes with a.
7. Let R be a ring in which $x^4 = x$ for every $x \in R$. Prove that R is commutative.
8. If F is a finite field, show that:
   (a) There exists a prime p such that $pa = 0$ for all $a \in F$.
   (b) If F has q elements, then $q = p^n$ for some integer n. (**Hint:** Cauchy's Theorem)
9. Let p be an odd prime and let $1 + \frac{1}{2} + \cdots + 1/(p-1) = a/b$, where a, b are integers. Show that $p \mid a$. (**Hint:** As a runs through U_p, so does a^{-1}.)
10. If p is a prime and $p > 3$, show that if $1 + \frac{1}{2} + \cdots + 1/(p-1) = a/b$, where a, b are integers, then $p^2 \mid a$. (**Hint:** Consider $1/a^2$ as a runs through U_p.)

3. IDEALS, HOMOMORPHISMS, AND QUOTIENT RINGS

In studying groups, it turned out that homomorphisms, and their kernels—the normal subgroups—played a central role. There is no reason to expect that the same thing should not be true for rings. As a matter of fact, the analogs, in the setting of rings, of homomorphism and normal subgroup do play a key role.

With the background we have acquired about such things in group theory, the parallel development for rings should be easy and quick. And it will be! Without any further fuss we make the

Definition. The mapping $\varphi: R \to R'$ of the ring R into the ring R' is a *homomorphism* if

(a) $\varphi(a + b) = \varphi(a) + \varphi(b)$ and
(b) $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in R$.

Since a ring has two operations, it is only natural and just that we demand that both these operations be preserved under what we would call a ring homomorphism. *Furthermore, Property (a) in the Definition tells us that φ is a homomorphism of R viewed merely as an abelian group under $+$* into R' (also viewed as a group under its addition). So we can call on, and expect, certain results from this fact alone.

Just as we saw in Chapter 2, Section 5 for groups, the image of R under a homomorphism from R to R', is a subring of R', as defined in Chapter 4, Section 1 (Prove!).

Let $\varphi: R \to R'$ be a ring homomorphism and let $\text{Ker}\,\varphi = \{x \in R \mid \varphi(x) = 0\}$, the 0 being that of R'. What properties does $\text{Ker}\,\varphi$ enjoy? Clearly, from group theory $\text{Ker}\,\varphi$ is an additive subgroup of R. But much more is true. If $k \in \text{Ker}\,\varphi$ and $r \in R$, then $\varphi(k) = 0$, so $\varphi(kr) = \varphi(k)\varphi(r) = 0\varphi(r) = 0$, and similarly, $\varphi(rk) = 0$. So $\text{Ker}\,\varphi$ swallows up multiplication from the left and the right by arbitrary ring elements.

This property of $\text{Ker}\,\varphi$ is now abstracted to define the important analog in ring theory of the notion of normal subgroup in group theory.

Definition. Let R be a ring. A nonempty subset I of R is called an *ideal* of R if:

(a) I is an additive subgroup of R.
(b) Given $r \in R$, $a \in I$, then $ra \in I$ and $ar \in I$.

We shall soon see some examples of homomorphisms and ideals. But first we note that Part (b) in the definition of ideal really has a left and a right part. We could split it and define a set L of R to be a *left ideal* of R if L is an additive subgroup of R and given $r \in R$, $a \in L$, then $ra \in L$. So we require only left swallowing-up for a left ideal. We can similarly define *right ideals*. An ideal as we defined it is both a left and a right ideal of R. By all rights we should then call what we called an ideal a *two-sided ideal* of R. Indeed, in working in noncommutative ring theory one uses this name; *here, by "ideal" we shall always mean a two-sided ideal.* Except for some of the problems, we shall not use the notion of one-sided ideals in this chapter.

Before going on, we record what was done above for $\text{Ker}\,\varphi$ as

Lemma 4.3.1. If $\varphi: R \to R'$ is a homomorphism, then $\text{Ker}\,\varphi$ is an ideal of R.

We shall soon see that every ideal can be made the kernel of a homomorphism. Shades of what happens for normal subgroups of groups!

Finally, let K be an ideal of R. Since K is an additive subgroup of R, the quotient group R/K exists; it is merely the set of all cosets $a + K$ as a runs over R. But R is not just a group; it is, after all, a ring. Nor is K merely an additive subgroup of R; it is more than that, namely an ideal of R. We should be able to put all this together to make of R/K a ring.

How should we define a product in R/K in a natural way? What do we want to declare $(a + K)(b + K)$ to be? The reasonable thing is to define $(a + K)(b + K) = ab + K$, which we do. As always, the first thing that comes up is to show that this product is well-defined. Is it? We must show that if $a + K = a' + K$ and $b + K = b' + K$, then $(a + K)(b + K) = ab + K = a'b' + K = (a' + K)(b' + K)$. However, if $a + K = a' + K$, then $a - a' \in K$, so $(a - a')b \in K$, since K is an ideal of R (in fact, so far, since K is a right ideal of R). Because $b + K = b' + K$, we have $b - b' \in K$, so $a'(b - b') \in K$, since K is an ideal of R (in fact, since K is a left ideal of R). So both $(a - a')b = ab - a'b$ and $a'(b - b') = a'b - a'b'$ are in K. Thus $(ab - a'b) + (a'b - a'b') = ab - a'b' \in K$.

But this tells us (just from group theory) that $ab + K = a'b' + K$, exactly what we needed to have the product well-defined.

So R/K is now endowed with a sum and a product. Furthermore, the mapping $\varphi: R \to R/K$ defined by $\varphi(a) = a + K$ for $a \in R$ is a homomorphism of R onto R/K with kernel K. (Prove!) This tells us right away that R/K is a ring, being the homomorphic image of the ring R.

We summarize all this in

Theorem 4.3.2. Let K be an ideal of R. Then the quotient group R/K as an additive group is a ring under the multiplication $(a + K)(b + K) = ab + K$. Furthermore, the map $\varphi: R \to R/K$ defined by $\varphi(a) = a + K$ for $a \in R$ is a homomorphism of R onto R/K having K as its kernel. So R/K is a homomorphic image of R.

Just from group-theoretic consideration of R as an additive group, we have that if φ is a homomorphism of R into R', then it is 1-1 if and only if $\operatorname{Ker} \varphi = (0)$. As in groups, we define a homomorphism to be a *monomorphism* if it is 1-1. A monomorphism which is also onto is called an *isomorphism*. We define R and R' to be *isomorphic* if there is an *isomorphism* of R *onto* R'.

An *isomorphism* from a ring R onto itself is called an *automorphism* of R. For example, suppose R is the field $\mathbb{C}$ of complex numbers. Then the mapping from $\mathbb{C}$ to $\mathbb{C}$ sending each element of $\mathbb{C}$ to its complex conjugate is an automorphism of $\mathbb{C}$. (Prove!)

One would have to be an awful pessimist to expect that the homomorphism theorems proved in Section 7 of Chapter 2 fail in this setting. In fact they do hold, with the slightly obvious adaptation needed to make the proofs go through. We state the homomorphism theorems without any further ado, leaving the few details needed to complete the proofs to the reader.

Theorem 4.3.3 (First Homomorphism Theorem). Let the mapping $\varphi: R \to R'$ be a homomorphism of R onto R' with kernel K. Then $R' \simeq R/K$; in fact, the mapping $\psi: R/K \to R'$ defined by $\psi(a + K) = \varphi(a)$ defines an isomorphism of R/K onto R'.

Theorem 4.3.4 (Correspondence Theorem). Let the mapping $\varphi: R \to R'$ be a homomorphism of R onto R' with kernel K. If I' is an ideal of R', let $I = \{a \in R \mid \varphi(a) \in I'\}$. Then I is an ideal of R, $I \supset K$ and $I/K \simeq I'$. This sets up a 1-1 correspondence between all the ideals of R' and those ideals of R that contain K.

Theorem 4.3.5 (Second Homomorphism Theorem). Let A be a subring of a ring R and I an ideal of R. Then $A + I = \{a + i \mid a \in A, i \in I\}$ is a subring of R, I is an ideal of $A + I$, and $(A + I)/I \simeq A/(A \cap I)$.

Theorem 4.3.6 (Third Homomorphism Theorem). Let the mapping $\varphi: R \to R'$ be a homomorphism of R onto R' with kernel K. If I' is an ideal of R' and $I = \{a \in R \mid \varphi(a) \in I'\}$, then $R/I \simeq R'/I'$. Equivalently, if K is an ideal of R and $I \supset K$ is an ideal of R, then $R/I \simeq (R/K)/(I/K)$.

We close this section with an inspection of some of the things we have discussed in some examples.

Examples

1. As usual we use $\mathbb{Z}$, the ring of integers, for our first example. Let $n > 1$ be a fixed integer and let I_n be the set of all multiples of n; then I_n is an ideal of $\mathbb{Z}$. If $\mathbb{Z}_n$ is the integers mod n, define $\varphi: \mathbb{Z} \to \mathbb{Z}_n$ by $\varphi(a) = [a]$. As is easily seen, φ is a homomorphism of $\mathbb{Z}$ onto $\mathbb{Z}_n$ with kernel I_n. So by Theorem 4.3.3, $\mathbb{Z}_n \simeq \mathbb{Z}/I_n$. (This should come as no surprise, for that is how we originally introduced $\mathbb{Z}_n$.)

2. Let F be a field; what can the ideals of F be? Suppose that $I \neq (0)$ is an ideal of F; let $a \neq 0 \in I$. Then, since I is an ideal of F, $1 = a^{-1}a \in I$; but now, since $1 \in I$, $r1 = r \in I$ for every $r \in F$. In short, $I = F$. So *F has only the trivial ideals* (0) *and F itself.*

3. Let R be the ring of all rational numbers having odd denominators in their reduced form. Let I be those elements of R which in reduced form have an

even numerator; it is easy to see that I is an ideal of R. Define $\varphi: R \to \mathbb{Z}_2$, the integers mod 2, by $\varphi(a/b) = 0$ if a is even (a, b have no common factor) and $\varphi(a/b) = 1$ if a is odd. We leave it to the reader to verify that φ is a homomorphism of R onto $\mathbb{Z}_2$ with kernel I. Thus $\mathbb{Z}_2 \simeq R/I$. Give the explicit isomorphism of R/I onto $\mathbb{Z}_2$.

4. Let R be the ring of all rational numbers whose denominators (when in reduced form) are not divisible by p, p a fixed prime. Let I be those elements in R whose numerator is divisible by p; I is an ideal of R and $R/I \simeq \mathbb{Z}_p$, the integers mod p. (Prove!)

5. Let R be the ring of all real-valued continuous functions on the closed unit interval where $(f + g)(x) = f(x) + g(x)$ and $(fg)(x) = f(x)g(x)$ for $f, g \in R$, $x \in [0, 1]$. Let $I = \{f \in R \mid f(\frac{1}{2}) = 0\}$. We claim that I is an ideal of R. Clearly, it is an additive subgroup. Furthermore if $f \in I$ and $g \in R$, then $f(\frac{1}{2}) = 0$, so $(fg)(\frac{1}{2}) = f(\frac{1}{2})g(\frac{1}{2}) = 0g(\frac{1}{2}) = 0$. Thus $fg \in I$; since I is commutative, gf is also in I. So I is an ideal of R.

What is R/I? Given any $f \in R$, then

$$f(x) = (f(x) - f(\tfrac{1}{2})) + f(\tfrac{1}{2}) = g(x) + f(\tfrac{1}{2}),$$

where $g(x) = f(x) - f(\frac{1}{2})$. Because $g(\frac{1}{2}) = f(\frac{1}{2}) - f(\frac{1}{2}) = 0$, g is in I. So $g + I = I$. Thus $f + I = (f(\frac{1}{2}) + g) + I = f(\frac{1}{2}) + I$. Because $f(\frac{1}{2})$ is just a real number, R/I consists of the cosets $\alpha + I$ for α real. We claim that every real α comes up. For if $f(\frac{1}{2}) = \beta \neq 0$, then

$$\begin{aligned}\alpha\beta^{-1}f + I &= (\alpha\beta^{-1} + I)(f + I) = (\alpha\beta^{-1} + I)(f(\tfrac{1}{2}) + I) \\ &= (\alpha\beta^{-1} + 1)(\beta + I) = \alpha\beta^{-1}\beta + I = \alpha + I.\end{aligned}$$

So R/I consists of all $\alpha + I$, α real. Thus we have shown directly that R/I is isomorphic to the real field.

We now use Theorem 4.3.3 to show that $R/I \simeq$ real field. Let $\varphi: R \to \mathbb{R}$ be defined by $\varphi(f) = f(\frac{1}{2})$. Then it is easy to verify that φ is a homomorphism, φ is onto and Ker $\varphi = \{f \in R \mid f(\frac{1}{2}) = 0\}$; in other words, Ker $\varphi = I$. So $R/I \simeq$ image of $\varphi = \mathbb{R}$.

6. Let R be the integral quaternions, in other words, $R = \{\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k \mid \alpha_0, \alpha_1, \alpha_2, \alpha_3 \in \mathbb{Z}\}$. For a fixed prime p, let

$$I_p = \{\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k \in R \mid p \mid \alpha_i \text{ for } i = 0, 1, 2, 3\}.$$

The reader should verify that I_p is an ideal of R and that $R/I_p \simeq H(\mathbb{Z}_p)$ (see Problem 36 of Section 1 and the paragraph before it).

7. Let $R = \left\{ \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \middle| \, a, b \in \mathbb{R} \right\}$; R is a subring of the 2×2 matrices over the reals.

Let $I = \left\{ \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \middle| \, b \in \mathbb{R} \right\}$; I is easily seen to be an additive subgroup of R. Is it an ideal of R? Consider

$$\begin{pmatrix} x & y \\ 0 & x \end{pmatrix} \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & xb \\ 0 & 0 \end{pmatrix};$$

so it is in I. Similarly,

$$\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x & y \\ 0 & x \end{pmatrix} = \begin{pmatrix} 0 & bx \\ 0 & 0 \end{pmatrix},$$

so it, too, is in I. So I is an ideal of R. What is R/I? We approach it from two points of view.

Given $\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \in R$,

then $\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}$ so that

$$\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} + I = \left(\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \right) + I = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + I,$$

since $\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}$ is in I. Thus all the cosets of I in R look like $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + I$. If we map this onto a, that is, $\psi \left(\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} + I \right) = a$, we can check that ψ is an isomorphism onto the real field. So $R/I \simeq \mathbb{R}$.

We see that $R/I \simeq \mathbb{R}$ another way. Define $\varphi : R \to \mathbb{R}$ by $\varphi \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} = a$. We claim that φ is a homomorphism. For, given $\begin{pmatrix} a & b \\ 0 & a \end{pmatrix}, \begin{pmatrix} c & d \\ 0 & c \end{pmatrix}$, then

$$\varphi \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} = a, \quad \varphi \begin{pmatrix} c & d \\ 0 & c \end{pmatrix} = c,$$

$$\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} + \begin{pmatrix} c & d \\ 0 & c \end{pmatrix} = \begin{pmatrix} a + c & b + d \\ 0 & a + c \end{pmatrix},$$

and $\begin{pmatrix} a & b \\ 0 & a \end{pmatrix}\begin{pmatrix} c & d \\ 0 & c \end{pmatrix} = \begin{pmatrix} ac & ad + bc \\ 0 & ac \end{pmatrix}$; hence

$$\varphi\left(\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} + \begin{pmatrix} c & d \\ 0 & c \end{pmatrix}\right) = \varphi\begin{pmatrix} a + c & b + d \\ 0 & a + c \end{pmatrix}$$

$$= a + c = \varphi\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} + \varphi\begin{pmatrix} c & d \\ 0 & c \end{pmatrix}$$

and $$\varphi\left(\begin{pmatrix} a & b \\ 0 & a \end{pmatrix}\begin{pmatrix} c & d \\ 0 & c \end{pmatrix}\right) = \varphi\begin{pmatrix} ac & ad + bc \\ 0 & ac \end{pmatrix}$$

$$= ac = \varphi\begin{pmatrix} a & b \\ 0 & a \end{pmatrix}\varphi\begin{pmatrix} c & d \\ 0 & c \end{pmatrix}.$$

So φ is indeed a homomorphism of R onto $\mathbb{R}$. What is Ker φ? If $\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \in$ Ker φ, then $\varphi\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} = a$, but also $\varphi\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} = 0$, since $\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \in$ Ker φ.

Thus $a = 0$. From this we see that $I =$ Ker φ. So $R/I \simeq$ image of $\varphi = \mathbb{R}$ by Theorem 4.3.3.

8. Let $R = \left\{\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \middle| a, b \in \mathbb{R}\right\}$ and let $\mathbb{C}$ be the field of complex numbers. Define $\psi: R \to \mathbb{C}$ by $\psi\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = a + bi$. We leave it to the reader to verify that ψ is an isomorphism of R onto $\mathbb{C}$. So R is isomorphic to the field of complex numbers.

9. Let R be any commutative ring with 1. If $a \in R$, let $(a) = \{xa \mid x \in R\}$. We claim that (a) is an ideal of R. To see this, suppose that $u, v \in (a)$; thus $u = xa$, $v = ya$ for some $x, y \in R$, whence

$$u \pm v = xa \pm ya = (x \pm y)\, a \in (a).$$

Also, if $u \in (a)$ and $r \in R$, then $u = xa$, hence $ru = r(xa) = (rx)a$, so is in (a). Thus (a) is an ideal of R.

Note that if R is not commutative, then (a) need not be an ideal; but it is certainly a left ideal of R.

PROBLEMS

Easier Problems

1. If R is a commutative ring and $a \in R$, let $L(a) = \{x \in R \mid xa = 0\}$. Prove that $L(a)$ is an ideal of R.
2. If R is a commutative ring with 1 and R has no ideals other than (0) and itself, prove that R is a field. (**Hint:** Look at Example 9.)
*3. If $\varphi: R \to R'$ is a homomorphism of R *onto* R' and R has a unit element, 1, show that $\varphi(1)$ is the unit element of R'.
4. If I, J are ideals of R, define $I + J$ by $I + J = \{i + j \mid i \in I, j \in J\}$. Prove that $I + J$ is an ideal of R.
5. If I is an ideal of R and A is a subring of R, show that $I \cap A$ is an ideal of A.
6. If I, J are ideals of R, show that $I \cap J$ is an ideal of R.
7. Give a complete proof of Theorem 4.3.2.
8. Give a complete proof of Theorem 4.3.4.
9. Let $\varphi: R \to R'$ be a homomorphism of R onto R' with kernel K. If A' is a subring of R', let $A = \{a \in R \mid \varphi(a) \in A'\}$. Show that:
   (a) A is a subring of R, $A \supset K$.
   (b) $A/K \simeq A'$.
   (c) If A' is a left ideal of R', then A is a left ideal of R.
10. Prove Theorem 4.3.6.
11. In Example 3, give the explicit isomorphism of R/I onto $\mathbb{Z}_2$.
12. In Example 4, show that $R/I \simeq \mathbb{Z}_p$.
13. In Example 6, show that $R/I_p \simeq H(\mathbb{Z}_p)$.
14. In Example 8, verify that the mapping ψ given is an isomorphism of R onto $\mathbb{C}$.
15. If I, J are ideals of R, let IJ be the set of all *sums of elements of the form* ij, where $i \in I, j \in J$. Prove that IJ is an ideal of R.
16. Show that the ring of 2×2 matrices over the reals has nontrivial left ideals (and also nontrivial right ideals).
17. Prove Theorem 4.3.5.

If R, S are rings, define the *direct sum* of R and S, $R \oplus S$, by

$$R \oplus S = \{(r, s) \mid r \in R, s \in S\}$$

where $(r, s) = (r_1, s_1)$ if and only if $r = r_1, s = s_1$, and where

$$(r, s) + (t, u) = (r + t, s + u), \quad (r, s)(t, u) = (rt, su).$$

18. Show that $R \oplus S$ is a ring and that the subrings $\{(r, 0) \mid r \in R\}$ and $\{(0, s) \mid s \in S\}$ are ideals of $R \oplus S$ isomorphic to R and S, respectively.

19. If $R = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \middle| \, a, b, c \text{ real} \right\}$ and $I = \left\{ \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \middle| \, b \text{ real} \right\}$, show that:

(a) R is a ring.

(b) I is an ideal of R.

(c) $R/I \simeq F \oplus F$, where F is the field of real numbers.

20. If I, J are ideals of R, let $R_1 = R/I$ and $R_2 = R/J$. Show that $\varphi : R \to R_1 \oplus R_2$ defined by $\varphi(r) = (r + I, r + J)$ is a homomorphism of R into $R_1 \oplus R_2$ such that $\operatorname{Ker} \varphi = I \cap J$.

21. Let $\mathbb{Z}_{15}$ be the ring of integers mod 15. Show that $\mathbb{Z}_{15} \simeq \mathbb{Z}_3 \oplus \mathbb{Z}_5$.

Middle-Level Problems

22. Let $\mathbb{Z}$ be the ring of integers and m, n two relatively prime integers, I_m the multiples of m in $\mathbb{Z}$, and I_n the multiples of n in $\mathbb{Z}$.

(a) What is $I_m \cap I_n$?

(b) Use the result of Problem 20 to show that there is a one-to-one homomorphism from $\mathbb{Z}/I_{mn}$ to $\mathbb{Z}/I_m \oplus \mathbb{Z}/I_n$.

23. If m, n are relatively prime, prove that $\mathbb{Z}_{mn} \simeq \mathbb{Z}_m \oplus \mathbb{Z}_n$. (**Hint:** Use a counting argument to show that the homomorphism of Problem 22(b) is onto.)

**24.* Use the result of Problem 23 to prove the *Chinese Remainder Theorem*, which asserts that if m and n are relatively prime integers and a, b any integers, we can find an integer x such that $x \equiv a \bmod m$ and $x \equiv b \bmod n$ simultaneously.

25. Let R be the ring of 2×2 matrices over the real numbers; suppose that I is an ideal of R. Show that $I = (0)$ or $I = R$. (Contrast this with the result of Problem 16.)

Harder Problems

26. Let R be a ring with 1 and let S be the ring of 2×2 matrices over R. If I is an ideal of S show that there is an ideal J of R such that I consists of all the 2×2 matrices over J.

27. If $p_1, p_2, \ldots, p_n$ are distinct odd primes, show that there are exactly 2^n solutions of $x^2 \equiv x \bmod(p_1 \cdots p_n)$, where $0 \leq x < p_1 \cdots p_n$.

28. Suppose that R is a ring whose only left ideals are (0) and R itself. Prove that either R is a division ring or R has p elements, p a prime, and $ab = 0$ for every $a, b \in R$.

29. Let R be a ring with 1. An element $a \in R$ is said to have a *left inverse* if $ba = 1$ for some $b \in R$. Show that if the left inverse b of a is unique, then $ab = 1$ (so b is also a right inverse of a).

4. MAXIMAL IDEALS

This will be a section with one major theorem. The importance of this result will only become fully transparent when we discuss fields in Chapter 5. However, it is a result that stands on its own two feet. It isn't difficult to prove, but in mathematics the correlation between difficult and important isn't always that high. There are many difficult results that are of very little interest and of even less importance, and some easy results that are crucial. Of course, there are some results—many, many—which are of incredible difficulty and importance.

Lemma 4.4.1. Let R be a commutative ring with unit whose only ideals are (0) and itself. Then R is a field.

Proof. Let $a \neq 0$ be in R. Then $(a) = \{xa \mid x \in R\}$ is an ideal of R, as we verified in Example 9 in the preceding section. Since $a = 1a \in (a)$, $(a) \neq (0)$. Thus, by our hypothesis on R, $(a) = R$. But then, by the definition of (a), every element $i \in R$ is a multiple xa of a for some $x \in R$. In particular, because $1 \in R$, $1 = ba$ for some $b \in R$. This shows that a has the inverse b in R. So R is a field. □

In Theorem 4.3.4—the Correspondence Theorem—we saw that if $\varphi : R \to R'$ is a homomorphism of R onto R' with kernel K, then there is a 1-1 correspondence between ideals of R' and ideals of R that contain K. Suppose that there are no ideals other than K itself and R which contain K. What does this imply about R'? Since (0) in R' corresponds to K in R, and R' corresponds to all of R, we must conclude that in this case R' has no ideals other than (0) and itself. So if R' is commutative and has a unit element, then, by Lemma 4.4.1, R' must be a field.

This prompts the following definition.

Definition. A proper ideal M of R is a *maximal ideal* of R if the only ideals of R that contain M are M itself and R.

The discussion preceding this definition has already almost proved for us

Theorem 4.4.2. Let R be a commutative ring with 1, and let M be a maximal ideal of R. Then R/M is a field.

Proof. There is a homomorphism of R onto $R' = R/M$, and since $1 \in R$ we have that R' has $1 + M$ as its unit element. (See Problem 3, Section 3). Because M is a maximal ideal of R, we saw in the discussion above that R' has no nontrivial ideals. Thus, by Lemma 4.4.1, $R' = R/M$ is a field. □

This theorem will be our entry into the discussion of fields, for it will enable us to construct particularly desirable fields whenever we shall need them.

Theorem 4.4.2 has a converse. This is

Theorem 4.4.3. If R is a commutative ring with 1 and M an ideal of R such that R/M is a field, then M is a maximal ideal of R.

Proof. We saw in Example 2 of Section 3 that the only ideals in a field F are (0) and F itself. Since R/M is a field, it has only (0) and itself as ideals. But then, by the correspondence given us by Theorem 4.3.4, there can be no ideal of R property between M and R. Thus M is a maximal ideal of R. □

We give a few examples of maximal ideals in commutative rings.

Examples

1. Let $\mathbb{Z}$ be the integers and M an ideal of $\mathbb{Z}$. As an ideal of $\mathbb{Z}$ we certainly have that M is an additive subgroup of $\mathbb{Z}$, so must consist of all multiples of some fixed integer n. Thus since $R/M \simeq \mathbb{Z}_n$ and since $\mathbb{Z}_n$ is a field if and only if n is a prime, we see that M is a maximal ideal of $\mathbb{Z}$ if and only if M consists of all the multiples of some prime p. Thus the set of maximal ideals in $\mathbb{Z}$ corresponds to the set of prime numbers.

2. Let $\mathbb{Z}$ be the integers, and let $R = \{a + bi \mid a, b \in \mathbb{Z}\}$, a subring of $\mathbb{C}$ $(i^2 = -1)$. Let M be the set of all $a + bi$ in R, where $3 \mid a$ and $3 \mid b$. We leave it to the reader to verify that M is an ideal of R.

We claim that M is a maximal ideal of R. For suppose that $N \supset M$ and $N \neq M$ is an ideal of R. So there is an element $r + si \in N$, where 3 does not divide r or 3 does not divide s. Therefore, $3 \nmid (r^2 + s^2)$. (Prove using congruences mod 3 !) But $t = r^2 + s^2 = (r + si)(r - si)$, so is in N, since $r + si \in N$ and N is an ideal of R. So N has an integer $t = r^2 + s^2$ *not* divisible by 3. Thus $ut + 3v = 1$ for some integers u, v; but $t \in N$, hence $ut \in N$ and $3 \in M \subset N$, so $3v \in N$. Therefore, $1 = ut + 3v \in N$. Therefore, $(a + bi)1 \in N$, since N is an ideal of R, for all $a + bi \in R$. This tells us that $N = R$. So the only ideal of R above M is R itself. Consequently, M is a maximal ideal of R.

By Theorem 4.4.2 we know that R/M is a field. It can be shown (see Problem 2) that R/M is a field having nine elements.

3. Let R be as in Example 2 and let $I = \{a + bi \mid 5|a \text{ and } 5|b\}$. We assert that I is *not* a maximal ideal of R.

In R we can factor $5 = (2 + i)(2 - i)$. Let $M = \{x(2 + i) \mid x \in R\}$. M is an ideal of R, and since $5 = (2 + i)(2 - i)$ is in M, we see that $I \subset M$. Clearly, $I \neq M$ for $2 + i \in M$ and is not in I because $5 \nmid 2$. So $I \neq M$. Can $M = R$? If so, then $(2 + i)(a + bi) = 1$ for some a, b. This gives $2a - b = 1$ and $2b + a = 0$; these two equations imply that $5a = 2$, so $a = \frac{2}{5}$, $b = -\frac{1}{5}$. But $\frac{2}{5} \notin \mathbb{Z}$, $-\frac{1}{5} \notin \mathbb{Z}$; the element $a + bi = \frac{2}{5} - \frac{1}{5}i$ is *not* in R. So $M \neq R$.

One can show, however, that M is a maximal ideal of R. (See Problem 3.)

4. Let $R = \{a + b\sqrt{2} \mid a, b \text{ integers}\}$, which is a subring of the real field under the sum and product of real numbers. That R is a ring follows from

$$(a + b\sqrt{2}) + (c + d\sqrt{2}) = (a + c) + (b + d)\sqrt{2}$$

and

$$(a + b\sqrt{2})(c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2}.$$

Let $M = \{a + b\sqrt{2} \in R \mid 5|a \text{ and } 5|b\}$. M is easily seen to be an ideal of R. We leave it to the reader to show that M is a maximal ideal of R and that R/M is a field having 25 elements.

5. Let R be the ring of all real-valued continuous functions on the closed unit interval $[0, 1]$. We showed in Example 5 of Section 3 that if $M = \{f \in R \mid f(\frac{1}{2}) = 0\}$, then M is an ideal of R and R/M is isomorphic to the real field. Thus, by Theorem 4.4.3, M is a maximal ideal of R.

Of course, if we let $M_\gamma = \{f \in R \mid f(\gamma) = 0\}$, where $\gamma \in [0, 1]$, then M_γ is also a maximal ideal. It can be shown that every maximal ideal in R is of the form M_γ for some $\gamma \in [0, 1]$, but to prove it we would require some results from real variable theory.

What this example shows is that the maximal ideals in R correspond to the points of $[0, 1]$.

PROBLEMS

1. If a, b are integers and $3 \nmid a$ or $3 \nmid b$, show that $3 \nmid (a^2 + b^2)$.

2. Show that in Example 2, R/M is a field having nine elements.

3. In Example 3, show that $M = \{x(2 + i) \mid x \in R\}$ is a maximal ideal of R.

4. In Example 3, show that $R/M \simeq \mathbb{Z}_5$.
5. In Example 3, show that $R/I \simeq \mathbb{Z}_5 \oplus \mathbb{Z}_5$.
6. In Example 4, show that M is a maximal ideal of R.
7. In Example 4, show that R/M is a field having 25 elements.
8. Using Example 2 as a model, construct a field having 49 elements.

We make a short excursion back to congruences mod p, where p is an *odd* prime. If a is an integer such that $p \nmid a$ and $x^2 \equiv a \bmod p$ has a solution x in $\mathbb{Z}$, we say that a is a *quadratic residue* mod p. Otherwise, a is said to be a *quadratic nonresidue* mod p.

9. Show that $(p-1)/2$ of the numbers $1, 2, \ldots, p-1$ are quadratic residues and $(p-1)/2$ are quadratic nonresidues mod p. **[Hint:** Show that $\{x^2 \mid x \neq 0 \in \mathbb{Z}_p\}$ forms a group of order $(p-1)/2$.]
10. Let $m > 0$ be in $\mathbb{Z}$, and suppose that m is *not* a square in $\mathbb{Z}$. Let $R = \{a + \sqrt{m}\, b \mid a, b \in \mathbb{Z}\}$. Prove that under the operations of sum and product of real numbers R is a ring.
11. If p is an odd prime, let us set $I_p = \{a + \sqrt{m}\, b \mid p|a \text{ and } p|b\}$, where $a + \sqrt{m}\, b \in R$, the ring in Problem 10. Show that I_p is an ideal of R.
12. If m is a *quadratic nonresidue* mod p, show that the ideal I_p in Problem 11 is a maximal ideal of R.
13. In Problem 12 show that R/I_p is a field having p^2 elements.

5. POLYNOMIAL RINGS

The material that we consider in this section involves the notion of polynomial and the set of all polynomials over a given field. We hope that most readers will have some familiarity with the notion of polynomial from their high school days and will have seen some of the things one does with polynomials: factoring them, looking for their roots, dividing one by another to get a remainder, and so on. The emphasis we shall give to the concept and algebraic object known as a polynomial ring will be in a quite different direction from that given in high school.

Be that as it may, what we shall strive to do here is to introduce the ring of polynomials over a field and show that this ring is amenable to a careful dissection that reveals its innermost structure. As we shall see, this ring is very well-behaved. The development should remind us of what was done for the ring of integers in Section 5 of Chapter 1. Thus we shall run into the analog of Euclid's algorithm, greatest common divisor, divisibility, and possibly most important, the appropriate analog of prime number. This will lead to

unique factorization of a general polynomial into these "prime polynomials," and to the nature of the ideals and the maximal ideals in this new setting.

But the polynomial ring enjoys one feature that the ring of integers did not: the notion of a root of a polynomial. The study of the nature of such roots—which will be done, for the most part, in the next chapter—constitutes a large and important part of the algebraic history of the past. It goes under the title *Theory of Equations*, and in its honorable past, a large variety of magnificent results have been obtained in this area. Hopefully, we shall see some of these as our development progresses.

With this sketchy outline of what we intend to do out of the way, we now get down to the nitty-gritty of doing it.

Let F be a field. *By the ring of polynomials in x over F*, which we shall always write as $F[x]$, we mean the set of all formal expressions $p(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + a_nx^n$, $n \geq 0$, where the a_i, the *coefficients* of the *polynomial* $p(x)$, are in F. We sometimes employ the alternative notation: $p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$. In $F[x]$ we define equality, sum, and product of two polynomials so as to make of $F[x]$ a commutative ring as follows:

1. ***Equality.*** We declare $p(x) = a_0 + a_1x + \cdots + a_nx^n$ and $q(x) = b_0 + b_1x + \cdots + b_nx^n$ to be *equal* if and only if their corresponding coefficients are equal, that is, if and only if $a_i = b_i$ for all $i \geq 0$.

We combine this definition of equality of polynomials $p(x)$ and $q(x)$ with the convention that if

$$q(x) = b_0 + b_1x + \cdots + b_nx^n$$

and if $b_{m+1} = \cdots = b_n = 0$, then we can drop the last $n - m$ terms and write $q(x)$ as

$$q(x) = b_0 + b_1x + \cdots + b_mx^m.$$

This convention is observed in the definition of addition that follows, where s is the larger of m and n and we add coefficients $a_{n+1} = \cdots = a_s = 0$ if $n < s$ or $b_{m+1} = \cdots = b_s = 0$ if $m < s$.

2. ***Addition.*** If $p(x) = a_0 + a_1x + \cdots + a_nx^n$ and $q(x) = b_0 + b_1x + \cdots + b_mx^m$, we declare $p(x) + q(x) = c_0 + c_1x + \cdots + c_sx^s$, where for each i, $c_i = a_i + b_i$.

So we add polynomials by adding their corresponding coefficients.

The definition of multiplication is a little more complicated. We define it loosely at first and then more precisely.

3. ***Multiplication.*** If $p(x) = a_0 + a_1x + \cdots + a_nx^n$ and $q(x) = b_0 + b_1x + \cdots + b_mx^m$, we declare $p(x)q(x) = c_0 + c_1x + \cdots + c_tx^t$, where the c_i are determined by multiplying the expression out formally, using the distributive laws and the rules of exponents $x^ux^v = x^{u+v}$, and collecting terms. More formally,

$$c_i = a_ib_0 + a_{i-1}b_1 + \cdots + a_1b_{i-1} + a_0b_i \quad \text{for every } i.$$

We illustrate these operations with a simple example, but first a notational device: If some coefficient is 0, we just omit that term. Thus we write $9 + 0x + 7x^2 + 0x^3 - 14x^4$ as $9 + 7x^2 - 14x^4$.

Let $p(x) = 1 + 3x^2$, $q(x) = 4 - 5x + 7x^2 - x^3$. Then $p(x) + q(x) = 5 - 5x + 10x^2 - x^3$ while

$$\begin{aligned} p(x)q(x) &= (1 + 3x^2)(4 - 5x + 7x^2 - x^3) \\ &= 4 - 5x + 7x^2 - x^3 + 3x^2(4 - 5x + 7x^2 - x^3) \\ &= 4 - 5x + 7x^2 - x^3 + 12x^2 - 15x^3 + 21x^4 - 3x^5 \\ &= 4 - 5x + 19x^2 - 16x^3 + 21x^4 - 3x^5. \end{aligned}$$

Try this product using the c_i as given above.

In some sense this definition of $F[x]$ is not a definition at all. We have indulged in some hand waving in it. But it will do. We could employ sequences to formally define $F[x]$ more precisely, but it would merely cloud what to most readers is well known.

The first remark that we make—and do not verify—is that $F[x]$ is a commutative ring. To go through the details of checking the axioms for a commutative ring is a straightforward but laborious task. However, it is important to note

Lemma 4.5.1. $F[x]$ is a commutative ring with unit.

Definition. If $p(x) = a_0 + a_1x + \cdots + a_nx^n$ and $a_n \neq 0$, then the *degree* of $p(x)$, denoted by $\deg p(x)$, is n.

So the degree of a polynomial $p(x)$ is the highest power of x that occurs in the expression for $p(x)$ with a nonzero coefficient. Thus $\deg(x - x^2 + x^4) = 4$, $\deg(7x) = 1$, $\deg 7 = 0$. (Note that this definition does not assign a degree to 0. It is, however, sometimes convenient to adopt the convention that the degree of 0 be $-\infty$, in which case many degree related results will hold in this extended context.) The polynomials of degree 0 and the polynomial 0 are called the *constants*; thus the set of constants can be identified with F.

The degree function on $F[x]$ will play a similar role to that played by

the size of the integers in $\mathbb{Z}$, in that it will provide us with a version of Euclid's Algorithm for $F[x]$.

One immediate and important property of the degree function is that it behaves well for products.

Lemma 4.5.2. If $p(x)$, $q(x)$ are nonzero elements of $F[x]$, then $\deg(p(x)q(x)) = \deg p(x) + \deg q(x)$.

Proof. Let $m = \deg p(x)$ and $n = \deg q(x)$; thus the polynomial $p(x) = a_0 + a_1x + \cdots + a_mx^m$, where $a_m \neq 0$, and the polynomial $q(x) = b_0 + b_1x + \cdots + b_nx^n$, where $b_n \neq 0$. The highest power of x that can occur in $p(x)q(x)$ is x^{m+n}, from our definition of the product. What is the coefficient of x^{m+n}? The only way that x^{m+n} can occur is from $(a_mx^m)(b_nx^n) = a_mb_nx^{m+n}$. So the coefficient of x^{m+n} in $p(x)q(x)$ is a_mb_n, which is not 0, since $a_m \neq 0$, $b_n \neq 0$. Thus $\deg(p(x)q(x)) = m + n = \deg p(x) + \deg q(x)$, as claimed in the lemma. □

One also has some information about $\deg(p(x) + q(x))$. This is

Lemma 4.5.3. If $p(x)$, $q(x) \in F[x]$ and $p(x) + q(x) \neq 0$, then $\deg(p(x) + q(x)) \leq \max(\deg p(x), \deg q(x))$.

We leave the proof of Lemma 4.5.3 to the reader. It will play no role in what is to come, whereas Lemma 4.5.2 will be important. We put it in so that the "+" should not feel slighted vis-à-vis the product.

An immediate consequence of Lemma 4.5.2 is

Lemma 4.5.4. $F[x]$ is an integral domain.

Proof. If $p(x) \neq 0$ and $q(x) \neq 0$, then $\deg p(x) \geq 0$, $\deg q(x) \geq 0$, so $\deg(p(x)q(x)) = \deg p(x) + \deg q(x) \geq 0$. Therefore, $p(x)q(x)$ has a degree, so cannot be 0 (which has no degree assigned to it). Thus $F[x]$ is an integral domain. □

One of the things that we were once forced to learn was to divide one polynomial by another. How did we do this? The process was called *long division*. We illustrate with an example how this was done, for what we do in the example is the model of what we shall do in the general case.

We want to divide $2x^2 + 1$ into $x^4 - 7x + 1$. We do it schematically as follows:

$$
\begin{array}{r}
\tfrac{1}{2}x^2 - \tfrac{1}{4} \\
(2x^2 + 1)\overline{)\,x^4 \qquad -7x + 1} \\
x^4 + \tfrac{1}{2}x^2 \\
\hline
-\tfrac{1}{2}x^2 - 7x + 1 \\
-\tfrac{1}{2}x^2 \qquad -\tfrac{1}{4} \\
\hline
-7x + 1\tfrac{1}{4}
\end{array}
$$

and we interpret this as:

$$x^4 - 7x + 1 = (2x^2 + 1)(\tfrac{1}{2}x^2 - \tfrac{1}{4}) + (-7x + \tfrac{5}{4})$$

and $-7x + \frac{5}{4}$ is called the *remainder* in this division.

What exactly did we do? First, where did the $\frac{1}{2}x^2$ come from? It came from the fact that when we multiply $2x^2 + 1$ by $\frac{1}{2}x^2$ we get x^4, the highest power occurring in $x^4 - 7x + 1$. So subtracting $\frac{1}{2}x^2(2x^2 + 1)$ from $x^4 - 7x + 1$ gets rid of the x^4 term and we go on to what is left and repeat the procedure.

This "repeat the procedure" suggests induction, and that is how we shall carry out the proof. But keep in mind that all we shall be doing is what we did in the example above.

What this gives us is something like Euclid's Algorithm, in the integers. However, here we call it the *Division Algorithm.*

Theorem 4.5.5 (Division Algorithm). Given the polynomials $f(x)$, $g(x) \in F[x]$, where $g(x) \neq 0$, then

$$f(x) = q(x)g(x) + r(x),$$

where $q(x), r(x) \in F[x]$ and $r(x) = 0$ or $\deg r(x) < \deg g(x)$.

Proof. We go by induction on $\deg f(x)$. If either $f(x) = 0$ or $\deg f(x) < \deg g(x)$, then $f(x) = 0g(x) + f(x)$, which satisfies the conclusion of the theorem.

So suppose that $\deg f(x) \geq \deg g(x)$; thus the polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$, where $a_m \neq 0$ and the polynomial $g(x) = b_0 + b_1x + \cdots + b_nx^n$, where $b_n \neq 0$ and where $m \geq n$.

Consider

$$\frac{a_m}{b_n}x^{m-n}g(x) = \frac{a_m}{b_n}x^{m-n}(b_0 + b_1x + \cdots + b_nx^n)$$

$$= \frac{a_mb_0}{b_n}x^{m-n} + \cdots + a_mx^m.$$

Thus $(a_m/b_n)x^{m-n}g(x)$ *has the same degree and same leading coefficient as does* $f(x)$, *so* $f(x) - (a_m/b_n)x^{m-n}g(x) = h(x)$ is such that the relation $\deg h(x) < \deg f(x)$ holds. Thus, by induction,

$$h(x) = q_1(x)g(x) + r(x), \quad \text{where} \quad q_1(x), r(x) \in F[x]$$

and $r(x) = 0$ or $\deg r(x) < \deg g(x)$. Remembering what $h(x)$ is, we get

$$h(x) = f(x) - \frac{a_m}{b_n}x^{m-n}g(x) = q_1(x)g(x) + r(x)$$

so

$$f(x) = \left(\frac{a_m}{b_n}x^{m-n} + q_1(x)\right)g(x) + r(x).$$

If $q(x) = (a_m/b_n)x^{m-n} + q_1(x)$, we have achieved the form claimed in the theorem. □

The Division Algorithm has one immediate application: It allows us to determine the nature of all the ideals of $F[x]$. As we see in the next theorem, an ideal of $F[x]$ must merely consist of all multiples, by elements of $F[x]$, of some *fixed* polynomial.

Theorem 4.5.6. If $I \neq (0)$ is an ideal of $F[x]$, then $I = \{f(x)g(x) \mid f(x) \in F[x]\}$; that is, I consists of all multiples of the fixed polynomial $g(x)$ by the elements of $F[x]$.

Proof. To prove the theorem, we need to produce that fixed polynomial $g(x)$. Where are we going to dig it up? The one control we have numerically on a given polynomial is its degree. So why not use the degree function as the mechanism for finding $g(x)$?

Since $I \neq (0)$ there are elements in I having nonnegative degree. So there is a polynomial $g(x) \neq 0$ in I of minimal degree; that is, $g(x) \neq 0$ is in I and if $0 \neq t(x) \in I$, then $\deg t(x) \geq \deg g(x)$. Thus, by the division algorithm, $t(x) = q(x)g(x) + r(x)$, where $r(x) = 0$ or $\deg r(x) < \deg g(x)$. But since $g(x) \in I$ and I is an ideal of $F[x]$, we have that $q(x)g(x) \in I$. By assumption, $t(x) \in I$, thus $t(x) - q(x)g(x)$ is in I, so $r(x) = t(x) - q(x)g(x)$ is in I. Since $g(x)$ has minimal degree for the elements of I and $r(x) \in I$, $\deg r(x)$ *cannot* be less than $\deg g(x)$. So we are left with $r(x) = 0$. But this says that $t(x) = q(x)g(x)$. So every element in I is a multiple of $g(x)$. On the other hand,

since $g(x) \in I$ and I is an ideal of $F[x]$, $f(x)g(x) \in I$ for all $f(x) \in F[x]$. The net result of all this is that $I = \{f(x)g(x) \mid f(x) \in F[x]\}$. □

Definition. An integral domain R is called a *principal ideal domain* if every ideal I in R is of the form $I = \{xa \mid x \in R\}$ for some $a \in I$.

Theorem 4.5.6 can be stated as: $F[x]$ *is a principal ideal domain.*

We shall write the *ideal generated by a given polynomial*, $g(x)$, namely $\{f(x)g(x) \mid f(x) \in F[x]\}$, as $(g(x))$.

The proof showed that if I is an ideal of $F[x]$, then $I = (g(x))$, where $g(x)$ is a polynomial of lowest degree contained in I. But $g(x)$ is not unique, for if $a \neq 0 \in F$, then $ag(x)$ is in I and has the same degree as $g(x)$, so $I = (ag(x))$.

To get some sort of uniqueness in all this, we single out a class of polynomials.

Definition. $f(x) \in F[x]$ is a *monic polynomial* if the coefficient of its highest power is 1.

Thus $f(x)$ is monic means that

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0.$$

We leave to the reader to show that if I is an ideal of $F[x]$, then there is *only one* monic polynomial of lowest degree in I. Singling this out as the generator of I does give us a "monic" uniqueness for the generation of I.

Our next step in this parallel development with what happens in the integers is to have the notion of one polynomial dividing another.

Definition. Suppose $f(x)$, $g(x) \in F[x]$, with $g(x) \neq 0$. We say that $g(x)$ divides $f(x)$, written as $g(x) \mid f(x)$, if $f(x) = a(x)g(x)$ for some $a(x) \in F[x]$.

Note that if $f(x) \neq 0$ and $g(x) \mid f(x)$, then $\deg g(x) \leq \deg f(x)$ by Lemma 4.5.2. Moreover, the ideals $(f(x))$ and $(g(x))$ of $F[x]$, generated by $f(x)$ and $g(x)$, respectively, satisfy the containing relation $(f(x)) \subset (g(x))$. (Prove!)

We again emphasize the parallelism between $\mathbb{Z}$, the integers, and $F[x]$ by turning to the notion of *greatest common divisor.* In order to get some sort of uniqueness, we shall insist that the greatest common divisor always be a monic polynomial.

Definition. For any two polynomials $f(x)$ and $g(x) \in F[x]$ (not both 0), the polynomial $d(x) \in F[x]$ is the *greatest common divisor* of $f(x)$, $g(x)$ if $d(x)$ is a monic polynomial such that:

(a) $d(x) \mid f(x)$ and $d(x) \mid g(x)$.

(b) If $h(x) \mid f(x)$ and $h(x) \mid g(x)$, then $h(x) \mid d(x)$.

Although we defined the greatest common divisor of two polynomials, we neither know, as yet, that it exists, nor what its form may be. We could define it in another, and equivalent, way as the *monic polynomial of highest degree that divides both $f(x)$ and $g(x)$*. If we did that, its existence would be automatic, but we would not know its form.

Theorem 4.5.7. Given $f(x)$ and $g(x) \neq 0$ in $F[x]$, then their greatest common divisor $d(x) \in F[x]$ exists; moreover, $d(x) = a(x)f(x) + b(x)g(x)$ for some $a(x), b(x) \in F[x]$.

Proof. Let I be the set of all $r(x)f(x) + s(x)g(x)$ as $r(x)$, $s(x)$ run freely over $F[x]$. We claim that I is an ideal of R. For,

$$(r_1(x)f(x) + s_1(x)g(x)) + (r_2(x)f(x) + s_2(x)g(x))$$
$$= (r_1(x) + r_2(x))f(x) + (s_1(x) + s_2(x))g(x),$$

so is again in I, and for $t(x) \in F[x]$,

$$t(x)(r(x)f(x) + s(x)g(x)) = (t(x)r(x))f(x) + (t(x)s(x))g(x),$$

so it, too, is again in I. Thus I is an ideal of $F[x]$. Since $g(x) \neq 0$, we know that $I \neq 0$, since both $f(x)$ and $g(x)$ are in I.

Since $I \neq 0$ is an ideal of $F[x]$, it is generated by a *unique monic* polynomial $d(x)$ (Theorem 4.5.6). Since $f(x)$, $g(x)$ are in I, they must then be multiples of $d(x)$ by elements of $F[x]$. This assures us that $d(x) \mid f(x)$ and $d(x) \mid g(x)$.

Because $d(x) \in I$ and I is the set of all $r(x)f(x) + s(x)g(x)$, we have that $d(x) = a(x)f(x) + b(x)g(x)$ for some appropriate $a(x)$, $b(x) \in F[x]$. Thus if $h(x) \mid f(x)$ and $h(x) \mid g(x)$, then $h(x) \mid (a(x)f(x) + b(x)g(x)) = d(x)$. So $d(x)$ is the greatest common divisor of $f(x)$ and $g(x)$.

This proves the theorem; the uniqueness of $d(x)$ is guaranteed by the demand that we have made that the greatest common divisor be monic. □

Another way to see the uniqueness of $d(x)$ is from

Lemma 4.5.8. If $f(x) \neq 0$, $g(x) \neq 0$ are in $F[x]$ and $f(x) \mid g(x)$ and $g(x) \mid f(x)$, then $f(x) = ag(x)$, where $a \in F$.

Proof. By the mutual divisibility condition on $f(x)$ and $g(x)$ we have, by Lemma 4.5.2, $\deg f(x) \leq \deg g(x) \leq \deg f(x)$, so $\deg f(x) = \deg g(x)$. But $f(x) = a(x)g(x)$, so

$$\deg f(x) = \deg a(x) + \deg g(x) = \deg a(x) + \deg f(x),$$

in consequence of which $\deg a(x) = 0$, so $a(x) = a$, an element of F. □

We leave the proof of the uniqueness of the greatest common divisor via Lemma 4.5.8 to the reader.

Definition. The polynomials $f(x)$, $g(x)$ in $F[x]$ are said to be *relatively prime* if their greatest common divisor is 1.

Although it is merely a very special case of Theorem 4.5.7, to emphasize it and to have it to refer to, we state:

Theorem 4.5.9. If $f(x)$, $g(x) \in F[x]$ are relatively prime, then $a(x)f(x) + b(x)g(x) = 1$ for some $a(x)$, $b(x) \in F[x]$. Conversely, if $a(x)f(x) + b(x)g(x) = 1$ for some $a(x)$, $b(x) \in F[x]$, then $f(x)$ and $g(x)$ are relatively prime.

Proof. We leave this "conversely" part to the reader as exercise. □

As with the integers, we have

Theorem 4.5.10. If $q(x)$ and $f(x)$ are relatively prime and if $q(x) \mid f(x)g(x)$, then $q(x) \mid g(x)$.

Proof. By Theorem 4.5.9 $a(x)f(x) + b(x)q(x) = 1$ for some $a(x)$, $b(x) \in F[x]$. Therefore,

$$a(x)f(x)g(x) + b(x)q(x)g(x) = g(x). \tag{1}$$

Since $q(x) \mid b(x)g(x)q(x)$ and $q(x) \mid f(x)g(x)$ by hypothesis, $q(x)$ divides the left-hand side of the relation in (1). Thus $q(x)$ divides the right-hand side of (1), that is, $q(x) \mid g(x)$, the desired conclusion. □

We are now ready to single out the important class of polynomials that will play the same role as prime objects in $F[x]$ as did the prime numbers in $\mathbb{Z}$.

Definition. The polynomial $p(x) \in F[x]$ is *irreducible* if $p(x)$ is of positive degree and given any polynomial $f(x)$ in $F[x]$, then either $p(x) \mid f(x)$ or $p(x)$ is relatively prime to $f(x)$.

We should note here that the definition implies that *a polynomial $p(x)$ of positive degree is irreducible in $F[x]$ if and only if $p(x)$ cannot be factored as a product of two polynomials of positive degree.* In other words, if $p(x) = a(x)b(x)$, where $a(x)$ and $b(x)$ are in $F[x]$, then either $a(x)$ is a constant (that is, an element of F), or $b(x)$ is constant. The proof of this fact is very similar to the proof of an analogous observation concerning two equivalent definitions of a prime number.

First suppose that $p(x)$ (of positive degree) cannot be factored as a product of two non–constant polynomials. Then, given any $f(x)$ in $F[x]$, we have only two possibilities for $(p(x), f(x))$, namely 1 or a monic polynomial of the form $c \cdot p(x)$, where c is an element of F. Thus $(p(x), f(x)) = 1$ or $p(x) \mid f(x)$, which shows that $p(x)$ is irreducible.

Now let $p(x)$ be irreducible in $F[x]$ and suppose $p(x) = a(x)b(x)$ for some $a(x)$, $b(x)$ in $F[x]$. According to the definition, we must have $p(x) \mid a(x)$ or $(p(x), a(x)) = 1$. If $p(x) \mid a(x)$, then $b(x)$ must be a constant. If, on the other hand, $p(x)$ and $a(x)$ are relatively prime, then by Theorem 4.5.10, $p(x) \mid b(x)$, and in this case $a(x)$ must be a constant. This shows that an irreducible polynomial cannot be factored as a product of two non–constant polynomials.

Note that the irreducibility of a polynomial depends on the field F. For instance, the polynomial $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$, where $\mathbb{Q}$ is the field of rational numbers, but $x^2 - 2$ is not irreducible in $\mathbb{R}[x]$, where $\mathbb{R}$ is the field of real numbers, for in $\mathbb{R}[x]$

$$x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2}).$$

Corollary to Theorem 4.5.10. If $p(x)$ is irreducible in $F[x]$ and $p(x) \mid a_1(x)a_2(x) \cdots a_k(x)$, where $a_1(x), \ldots, a_k(x)$ are in $F[x]$, then $p(x) \mid a_i(x)$ for some i.

Proof. We leave the proof to the reader. (See Theorem 1.5.6.) □

Aside from its other properties, an irreducible polynomial $p(x)$ in $F[x]$ enjoys the property that $(p(x))$, the ideal generated by $p(x)$ in $F[x]$, is a maximal ideal of $F[x]$. We prove this now.

Theorem 4.5.11. If $p(x) \in F[x]$, then the ideal $(p(x))$ generated by $p(x)$ in $F[x]$ is a maximal ideal of $F[x]$ if and only if $p(x)$ is irreducible in $F[x]$.

Proof. We first prove that if $p(x)$ is irreducible in $F[x]$, then the ideal $M = (p(x))$ is a maximal ideal of $F[x]$. For, suppose that N is an ideal of $F[x]$, and $N \supset M$. By Theorem 4.5.6,

$$N = (f(x)) \quad \text{for some} \quad f(x) \in F[x].$$

Because $p(x) \in M \subset N$, $p(x) = a(x)f(x)$, since every element in N is of this form. But $p(x)$ is irreducible in $F[x]$, hence $a(x)$ is a constant or $f(x)$ is a constant. If $a(x) = a \in F$, then $p(x) = af(x)$, so $f(x) = a^{-1}p(x)$. Thus $f(x) \in M$, which says that $N \subset M$, hence $N = M$. On the other hand, if $f(x) = b \in F$, then $1 = b^{-1}b \in N$, since N is an ideal of $F[x]$, thus $g(x)1 \in N$ for all $g(x) \in F[x]$. This says that $N = F[x]$. Therefore, we have shown M to be a maximal ideal of $F[x]$.

In the other direction, suppose that $M = (p(x))$ is a maximal ideal of $F[x]$. If $p(x)$ is not irreducible, then $p(x) = a(x)b(x)$, where $\deg a(x) \geq 1$, $\deg b(x) \geq 1$. Let $N = (a(x))$; then, since $p(x) = a(x)b(x)$, $p(x) \in N$. Therefore, $M \subset N$. Since $\deg a(x) \geq 1$, $N = (a(x)) \neq F[x]$, since every element in $(a(x))$ has degree at least that of $a(x)$. By the maximality of M we conclude that $M = N$. But then $a(x) \in N = M$, which tells us that $a(x) = f(x)p(x)$; combined with $p(x) = a(x)b(x) = b(x)f(x)p(x)$, we get that $b(x)f(x) = 1$. Since $\deg 1 = 0 < \deg b(x) \leq \deg(b(x)f(x)) = \deg 1 = 0$, we have reached a contradiction. Thus $p(x)$ is irreducible. □

This theorem is important because it tells us exactly what the maximal ideals of $F[x]$ are, namely the ideals generated by the irreducible polynomials. If M is a maximal ideal of $F[x]$, $F[x]/M$ is a field, and this field contains F (or more precisely, the field $\{a + M \mid a \in F\}$, which is isomorphic to F). This allows us to construct decent fields $K \supset F$, the decency of which lies in that $p(x)$ has a *root* in K. The exact statement and explanation of this we postpone until Chapter 5.

The last topic in this direction that we want to discuss is the factorization of a given polynomial as a product of irreducible ones. Note that if $p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$, $a_0 \neq 0$, is irreducible in $F[x]$, then so is $a_0^{-1}p(x)$ irreducible in $F[x]$; however, $a_0^{-1}p(x)$ has the advantage of being monic. So we have this monic irreducible polynomial trivially obtainable from $p(x)$ itself. This will allow us to make more precise the uniqueness part of the next theorem.

Theorem 4.5.12. Let $f(x) \in F[x]$ be of positive degree. Then either $f(x)$ is irreducible in $F[x]$ or $f(x)$ is the product of irreducible polynomials in $F[x]$. In fact, then,

$$f(x) = ap_1(x)^{m_1}p_2(x)^{m_2} \cdots p_k(x)^{m_k},$$

where a is the leading coefficient of $f(x)$, $p_1(x), \ldots, p_k(x)$ are monic and irreducible in $F[x]$, $m_1 > 0, \ldots, m_k > 0$ and this factorization in this form is unique up to the order of the $p_i(x)$.

Proof. We first show the first half of the theorem, namely that $f(x)$ is irreducible or the product of irreducibles. The proof is exactly the same as that of Theorem 1.5.7, with a slight, obvious adjustment.

We go by induction on $\deg f(x)$. If $\deg f(x) = 1$, then $f(x) = ax + b$ with $a \neq 0$ and is clearly irreducible in $F[x]$. So the result is true in this case.

Suppose, then, that the theorem is correct for all $a(x) \in F[x]$ such that $\deg a(x) < \deg f(x)$. If $f(x)$ is irreducible, then we have nothing to prove. Otherwise, $f(x) = a(x)b(x)$, $a(x)$ and $b(x) \in F[x]$ and $\deg a(x) < \deg f(x)$ and $\deg b(x) < \deg f(x)$. By the induction, $a(x)$ [and $b(x)$] is irreducible or is the product of irreducibles. But then $f(x)$ is the product of irreducible polynomials in $F[x]$. This completes the induction, and so proves the opening half of the theorem.

Now to the uniqueness half. Again we go by induction on $\deg f(x)$. If $\deg f(x) = 1$, then $f(x)$ is irreducible and the uniqueness is clear.

Suppose the result true for polynomials of degree less than $\deg f(x)$. Suppose that

$$f(x) = ap_1(x)^{m_1}p_2(x)^{m_2} \cdots p_k(x)^{m_k} = aq_1(x)^{n_1} \cdots q_r(x)^{n_r},$$

where the $p_i(x)$, $q_i(x)$ are monic irreducible polynomials and the m_i, n_i are all positive and a is the leading coefficient of $f(x)$, that is, the coefficient of the highest power term of $f(x)$. Since $p_1(x) \mid f(x)$, we have that $p_1(x) \mid q_1(x)^{n_1} \cdots q_r(x)^{n_r}$, so by the corollary to Theorem 4.5.10, $p_1(x) \mid q_i(x)$ for some i. Since $q_i(x)$ is monic and irreducible, as is $p_1(x)$, we get $p_1(x) = q_i(x)$. We can suppose (on renumbering) that $p_1(x) = q_1(x)$. Thus

$$\begin{aligned}\frac{f(x)}{p_1(x)} &= ap_1(x)^{m_1-1}p_2(x)^{m_2} \cdots p_k(x)^{m_k} \\ &= ap_1(x)^{n_1-1}q_2(x)^{n_2} \cdots q_r(x)^{n_r}.\end{aligned}$$

By induction we have unique factorization in the required form for $f(x)/p_1(x)$, whose degree is less than $\deg f(x)$. Hence we obtain that $m_1 - 1 = n_1 - 1$ (so $m_1 = n_1$), $m_2 = n_2, \ldots, m_k = n_k$, $r = k$ and $p_2(x) = q_2(x), \ldots, p_k(x) = q_k(x)$, on renumbering the q's appropriately. This completes the induction and proves the theorem. □

We have pointed out how similar the situation is for the integers $\mathbb{Z}$ and the polynomial ring $F[x]$. This suggests that there should be a wider class of rings, of which the two examples $\mathbb{Z}$ and $F[x]$ are special cases, for which much of the argumentation works. It worked for $\mathbb{Z}$ and $F[x]$ because we had a measure of size in them, either by the size of an integer or the degree of a polynomial. This measure of size was such that it allowed a Euclid-type algorithm to hold.

This leads us to define a class of rings, the *Euclidean rings*.

Definition. An integral domain R is a *Euclidean ring* if there is a function d from the *nonzero* elements of R to the nonnegative integers that satisfies:

(a) For $a \neq 0, b \neq 0 \in R$, $d(a) \leq d(ab)$.

(b) Given $a \neq 0, b \neq 0$, there exist q and $r \in R$ such that $b = qa + r$, where $r = 0$ or $d(r) < d(a)$.

The interested student should try to see which of the results proved for polynomial rings (and the integers) hold in a general Euclidean ring. Aside from a few problems involving Euclidean rings, we shall not go any further with this interesting class of rings.

The final comment we make here is that what we did for polynomials over a field we could *try* to do for polynomials over an arbitrary ring. That is, given any ring R (commutative or noncommutative), we could define the polynomial ring $R[x]$ in x over R by defining equality, addition, and multiplication exactly as we did in $F[x]$, for F a field. The ring so constructed, $R[x]$, is a very interesting ring, whose structure is tightly interwoven with that of R itself. It would be too much to expect that all, or even any, of the theorems proved in this section would carry over to $R[x]$ for a general ring R.

PROBLEMS

In the following problems, F will always denote a field.

Easier Problems

1. If F is a field, show that the only invertible elements in $F[x]$ are the nonzero elements of F.

2. If R is a ring, we introduce the ring $R[x]$ of polynomials in x over R, just as we did $F[x]$. Defining $\deg f(x)$ for $f(x) \in R[x]$ as we did in $F[x]$, show that:

(a) $\deg(f(x)g(x)) \leq \deg f(x) + \deg g(x)$ if $f(x)g(x) \neq 0$.

(b) There is a commutative ring R such that we can find $f(x)$, $g(x)$ in $R[x]$ with $\deg(f(x)g(x)) < \deg f(x) + \deg g(x)$.

3. Find the greatest common divisor of the following polynomials over $\mathbb{Q}$, the field of rational numbers.

(a) $x^3 - 6x + 7$ and $x + 4$.

(b) $x^2 - 1$ and $2x^7 - 4x^5 + 2$.

(c) $3x^2 + 1$ and $x^6 + x^4 + x + 1$.
(d) $x^3 - 1$ and $x^7 - x^4 + x^3 - 1$.

4. Prove Lemma 4.5.3.

5. In Problem 3, let $I = \{f(x)a(x) + g(x)b(x)\}$, where $f(x)$, $g(x)$ run over $\mathbb{Q}[x]$ and $a(x)$ is the first polynomial and $b(x)$ the second one in each part of the problem. Find $d(x)$, so that $I = (d(x))$ for Parts (a), (b), (c), and (d).

6. If $g(x), f(x) \in F[x]$ and $g(x) \mid f(x)$, show that $(f(x)) \subset (g(x))$.

7. Prove the uniqueness of the greatest common divisor of two polynomials in $F[x]$ by using Lemma 4.5.8.

8. If $f(x)$, $g(x) \in F[x]$ are relatively prime and $f(x) \mid h(x)$ and $g(x) \mid h(x)$, show that $f(x)g(x) \mid h(x)$.

9. Prove the Corollary to Theorem 4.5.10.

10. Show that the following polynomials are irreducible over the field F indicated.
(a) $x^2 + 7$ over F = real field = $\mathbb{R}$.
(b) $x^3 - 3x + 3$ over F = rational field = $\mathbb{Q}$.
(c) $x^2 + x + 1$ over $F = \mathbb{Z}_2$.
(d) $x^2 + 1$ over $F = \mathbb{Z}_{19}$.
(e) $x^3 - 9$ over $F = \mathbb{Z}_{13}$.
(f) $x^4 + 2x^2 + 2$ over $F = \mathbb{Q}$.

11. If $p(x) \in F[x]$ is of degree 3 and $p(x) = a_0x^3 + a_1x^2 + a_2x + a_3$, show that $p(x)$ if irreducible over F if there is no element $r \in F$ such that $p(r) = a_0r^3 + a_1r^2 + a_2r + a_3 = 0$.

12. If $F \subset K$ are two fields and $f(x)$, $g(x) \in F[x]$ are relatively prime in $F[x]$, show that they are relatively prime in $K[x]$.

Middle-Level Problems

13. Let $\mathbb{R}$ be the field of real numbers and $\mathbb{C}$ that of complex numbers. Show that $\mathbb{R}[x]/(x^2 + 1) \simeq \mathbb{C}$. [**Hint:** If $A = \mathbb{R}[x]/(x^2 + 1)$, let u be the image of x in A; show that every element in A is of the form $a + bu$, where $a, b \in \mathbb{R}$ and $u^2 = -1$.]

14. Let $F = \mathbb{Z}_{11}$, the integers mod 11.
(a) Let $p(x) = x^2 + 1$; show that $p(x)$ is irreducible in $F[x]$ and that $F[x]/(p(x))$ is a field having 121 elements.
(b) Let $p(x) = x^3 + x + 4 \in F[x]$; show that $p(x)$ is irreducible in $F[x]$ and that $F[x]/(p(x))$ is a field having 11^3 elements.

15. Let $F = \mathbb{Z}_p$ be the field of integers mod p, where p is a prime, and let $q(x) \in F[x]$ be irreducible of degree n. Show that $F[x]/(q(x))$ is a field having at most p^n elements. (See Problem 16 for a more exact statement.)

16. Let F, $q(x)$ be as in Problem 15; show that $F[x]/(q(x))$ has exactly p^n elements.

17. Let $p_1(x), p_2(x), \ldots, p_k(x) \in F[x]$ be distinct irreducible polynomials and let $q(x) = p_1(x)p_2(x) \cdots p_k(x)$. Show that

$$\frac{F[x]}{(q(x))} \simeq \frac{F[x]}{(p_1(x))} \oplus \frac{F[x]}{(p_2(x))} \oplus \cdots \oplus \frac{F[x]}{(p_k(x))}.$$

18. Let F be a finite field. Show that $F[x]$ contains irreducible polynomials of arbitrarily high degree. (**Hint:** Try to imitate Euclid's proof that there is an infinity of prime numbers.)

19. Construct a field having p^2 elements, for p an odd prime.

20. If R is a Euclidean ring, show that every ideal of R is principal.

21. If R is a Euclidean ring, show that R has a unit element.

22. If R is the ring of even integers, show that Euclid's algorithm is false in R by exhibiting two even integers for which the algorithm does not hold.

Harder Problems

23. Let $F = \mathbb{Z}_7$ and let $p(x) = x^3 - 2$ and $q(x) = x^3 + 2$ be in $F[x]$. Show that $p(x)$ and $q(x)$ are irreducible in $F[x]$ and that the fields $F[x]/(p(x))$ and $F[x]/(q(x))$ are isomorphic.

24. Let $\mathbb{Q}$ be the field of rational numbers, and let $q(x) = x^2 + x + 1$ be in $\mathbb{Q}(x)$. If α is a complex number such that $\alpha^2 + \alpha + 1 = 0$, show that the set $\{a + b\alpha \mid a, b \in \mathbb{Q}\}$ is a field in two ways; the first by showing it to be isomorphic to something you know is a field, the second by showing that if $a + ba \neq 0$, then its inverse is of the same form.

25. If p is a prime, show that $q(x) = 1 + x + x^2 + \cdots x^{p-1}$ is irreducible in $\mathbb{Q}[x]$.

26. Let R be a commutative ring in which $a^2 = 0$ only if $a = 0$. Show that if $q(x) \in R[x]$ is a zero-divisor in $R[x]$, then, if

$$q(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n,$$

there is an element $b \neq 0$ in R such that $ba_0 = ba_1 = \cdots = ba_n = 0$.

27. Let R be a ring and I an ideal of R. If $R[x]$ and $I[x]$ are the polynomial rings in x over R and I, respectively, show that:
(a) $I[x]$ is an ideal of $R[x]$.
(b) $R[x]/I[x] \simeq (R/I)[x]$.

Very Hard Problems

***28.** Do Problem 26 even if the condition "$a^2 = 0$ only if $a = 0$" does not hold in R.

29. Let $R = \{a + bi \mid a, b \text{ integers}\} \subset \mathbb{C}$. Let $d(a + bi) = a^2 + b^2$. Show that R is a Euclidean ring where d is its required Euclidean function. (R is known as the ring of *Gaussian integers* and plays an important role in number theory.)

6. POLYNOMIALS OVER THE RATIONALS

In our consideration of the polynomial ring $F[x]$ over a field F, the particular nature of F never entered the picture. All the results hold for arbitrary fields. However, there are results that exploit the explicit character of certain fields. One such field is that of the rational numbers.

We shall present two important theorems for $\mathbb{Q}[x]$, the polynomial ring over the rational field $\mathbb{Q}$. These results depend heavily on the fact that we are dealing with rational numbers. The first of these, *Gauss' Lemma*, relates the factorization over the rationals with factorization over the integers. The second one, known as the *Eisenstein Criterion*, gives us a method of constructing irreducible polynomials of arbitrary degree, at will, in $\mathbb{Q}[x]$. In this the field $\mathbb{Q}$ is highly particular. For instance, there is no easy algorithm for obtaining irreducible polynomials of arbitrary degree n over the field $\mathbb{Z}_p$ of the integers mod p, p a prime. Even over $\mathbb{Z}_2$ such an algorithm is nonexistent; it would be highly useful to have, especially for coding theory. But it just doesn't exist—so far.

We begin our consideration with two easy results.

Lemma 4.6.1. Let $f(x) \in \mathbb{Q}[x]$; then

$$f(x) = \frac{u}{m}(a_0x^n + a_1x^{n-1} + \cdots + a_n)$$

where $u, m, a_0, \ldots, a_n$ are integers and the $a_0, a_1, \ldots, a_n$ have no common factor greater than 1 (i.e., are relatively prime) and $(u, m) = 1$.

Proof. Since $f(x) \in \mathbb{Q}[x]$, $f(x) = q_0x^n + q_1x^{n-1} + \cdots + q_n$, where the q_i are rational numbers. So for $i = 0, 1, 2, \ldots, n$, $q_i = b_i/c_i$, where b_i, c_i are integers. Thus

$$f(x) = \frac{b_0}{c_0}x^n + \frac{b_1}{c_1}x^{n-1} + \cdots + \frac{b_n}{c_n};$$

clearing of denominators gives us

$$f(x) = \frac{1}{c_0c_1 \cdots c_n}(u_0x^n + u_1x^{n-1} + \cdots + u_n),$$

where the u_i are integers. If w is the greatest common divisor of $u_0, u_1, \ldots, u_n$, then each $u_i = wa_i$, where $a_0, a_1, \ldots, a_n$ are relatively prime integers. Then

$$f(x) = \frac{w}{c_0c_1c_2 \cdots c_n}(a_0x^n + a_1x^{n-1} + \cdots + a_n);$$

canceling out the greatest common factor of w and $c_0c_1 \cdots c_n$ gives us

$$f(x) = \frac{u}{m}(a_0x^n + \cdots + a_n),$$

where u, m are relatively prime integers, as is claimed in the lemma. □

The next result is a result about a particular homomorphic image of $R[x]$ for any ring R.

Lemma 4.6.2. If R is any ring and I an ideal of R, then $I[x]$, the polynomial ring in x over I, is an ideal of $R[x]$. Furthermore, $R[x]/I[x] \simeq (R/I)[x]$, the polynomial ring in x over R/I.

Proof. Let $\overline{R} = R/I$; then there is a homomorphism $\varphi : R \to \overline{R}$, defined by $\varphi(a) = a + I$, whose kernel is I. Define $\Phi : R[x] \to \overline{R}[x]$ by: If

$$f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n,$$

then

$$\Phi(f(x)) = \varphi(a_0)x^n + \varphi(a_1)x^{n-1} + \cdots + \varphi(a_n).$$

We leave it to the reader to prove that Φ is a homomorphism of $R[x]$ onto $\overline{R}[x]$. What is the kernel, Ker Φ, of Φ? If $f(x) = a_0x^n + \cdots + a_n$ is in Ker Φ, then $\Phi(f(x)) = 0$, the 0 element of $\overline{R}[x]$. Since

$$\Phi(f(x)) = \varphi(a_0)x^n + \varphi(a_1)x^{n-1} + \cdots + \varphi(a_n) = 0,$$

we conclude $\varphi(a_0) = 0$, $\varphi(a_1) = 0$, ... , $\varphi(a_n) = 0$, by the very definition of what we mean by the 0-polynomial in a polynomial ring. Thus each a_i is in

the kernel of φ, which happens to be I. Because $a_0, a_1, \ldots, a_n$ are in I, $f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$ is in $I[x]$. So Ker $\Phi \subset I[x]$. That $I[x] \subset$ Ker Φ is immediate from the definition of the mapping Φ. Hence $I[x] =$ Ker Φ. By the First Homomorphism Theorem (Theorem 4.3.3), the ring $I[x]$ is then an ideal of $R[x]$ and $\overline{R}[x] \simeq R[x]/\text{Ker } \Phi = R[x]/I[x]$. This proves the lemma, remembering that $\overline{R} = R/I$. □

As a very special case of the lemma we have the

Corollary. Let $\mathbb{Z}$ be the ring of integers, p a prime number in $\mathbb{Z}$, and $I = (p)$, the ideal of $\mathbb{Z}$ generated by p. Then $\mathbb{Z}[x]/I[x] \simeq \mathbb{Z}_p[x]$.

Proof. Since $\mathbb{Z}_p \simeq \mathbb{Z}/I$, the corollary follows by applying the lemma to $R = \mathbb{Z}$. □

We are ready to prove the first of the two major results we seek in this section.

Theorem 4.6.3 (Gauss' Lemma). If $f(x) \in \mathbb{Z}[x]$ is a monic polynomial and $f(x) = a(x)b(x)$, where $a(x)$ and $b(x)$ are in $\mathbb{Q}[x]$. Then $f(x) = a_1(x)b_1(x)$, where $a_1(x)$, $b_1(x)$ are monic polynomials in $\mathbb{Z}[x]$ and deg $a_1(x) =$ deg $a(x)$, deg $b_1(x) =$ deg $b(x)$.

Proof. Suppose that $f(x) \in \mathbb{Z}[x]$ and that $f(x) = a(x)b(x)$, where $a(x)$, $b(x) \in \mathbb{Q}[x]$, and deg $a(x) = s$, deg $b(x) = r$. By Lemma 4.6.1, we can express each of $a(x)$, $b(x)$ as a product of a rational number and a polynomial with integer coefficients. More precisely,

$$a(x) = \frac{u_1}{m_1}(a_0'x^s + a_1'x^{s-1} + \cdots + a_s') = \frac{u_1}{m_1}a_1(x),$$

where $a_0', a_1', \ldots, a_s'$ are relatively prime integers and

$$b(x) = \frac{u_2}{m_2}(b_0'x^r + b_1'x^{r-1} + \cdots + b_r') = \frac{u_2}{m_2}b_1(x),$$

where $b_0', b_1', \ldots, b_r'$ are relatively prime. Thus

$$f(x) = a(x)b(x) = \frac{u_1u_2}{m_1m_2}a_1(x)b_1(x) = \frac{v}{w}a_1(x)b_1(x),$$

where v and w are relatively prime, by canceling out the common factor of u_1u_2 and m_1m_2. Therefore, $wf(x) = va_1(x)b_1(x)$, and $f(x)$, $a_1(x)$, $b_1(x)$ *are all in* $\mathbb{Z}[x]$. Of course, we may assume with no loss of generality that the leading coefficients of $a_1(x)$ and $b_1(x)$ are positive.

If $w = 1$, then, since $f(x)$ is monic, we get that $va_0'b_0' = 1$ and this leads

easily to $v = 1$, $a_0' = b_0' = 1$ and so $f(x) = a_1(x)b_1(x)$, where both $a_1(x)$ and $b_1(x)$ are monic polynomials with integer coefficients. This is precisely the claim of the theorem, since $\deg a_1(x) = \deg a(x)$ and $\deg b_1(x) = \deg b(x)$.

Suppose then that $w \neq 1$; thus there is a prime p such that $p \mid w$ and, since $(v, w) = 1$, $p \nmid v$. Also, since the coefficients $a_0', a_1', \ldots, a_s'$ of $a_1(x)$ are relatively prime, there is an i such that $p \nmid a_i'$; similarly, there is a j such that $p \nmid b_j'$. Let $I = (p)$ be the ideal generated by p in $\mathbb{Z}$; then $\mathbb{Z}/I \simeq \mathbb{Z}_p$ and, by the Corollary to Lemma 4.6.2, $\mathbb{Z}[x]/I[x] \simeq \mathbb{Z}_p[x]$, so is an integral domain. However, since $p \mid w$, $\overline{w}$, the image of w in $\mathbb{Z}[x]/I[x]$, is 0, and since $p \nmid v$, $\overline{v}$ the image of v in $\mathbb{Z}[x]/I[x]$ is not 0. Thus $0\bar{f}(x) = \overline{v}\,\overline{a}_1(x)\overline{b}_1(x)$, where $\overline{v} \neq 0$ and $\overline{a}_1(x) \neq 0$, $\overline{b}_1(x) \neq 0$ because $p \nmid a_i'$ and $p \nmid b_j'$ for the given i, j above. This contradicts that $\mathbb{Z}[x]/I[x]$ is an integral domain. So $w \neq 1$ is not possible, and the theorem is proved. □

It might be instructive for the reader to try to show directly that if $x^3 + 6x - 7$ is the product of two polynomials having rational coefficients, then it is already the product of two monic polynomials with integer coefficients.

One should say something about C. F. Gauss (1777–1855), considered by many to be the greatest mathematician ever. His contributions in number theory, algebra, geometry, and so on, are of gigantic proportions. His contributions in physics and astronomy are also of such a great proportion that he is considered by physicists as one of their greats, and by the astronomers as one of the important astronomers of the past.

As we indicated at the beginning of this section, irreducible polynomials of degree n over a given field F may be very hard to come by. However, over the rationals, due to the next theorem, these exist in abundance and are very easy to construct.

Theorem 4.6.4 (The Eisenstein Criterion). Let $f(x) = x^n + a_1x^{n-1} + \cdots + a_n$ be a nonconstant polynomial with integer coefficients. Suppose that there is some prime p such that $p \mid a_1, p \mid a_2, \ldots, p \mid a_n$, but $p^2 \nmid a_n$. Then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Proof. Suppose that $f(x) = u(x)v(x)$, where $u(x)$, $v(x)$ *are of positive degree* and are polynomials in $\mathbb{Q}[x]$. By Gauss' Lemma we may assume that both $u(x)$ and $v(x)$ are monic polynomials with integer coefficients. Let $I = (p)$ be the ideal generated by p in $\mathbb{Z}$, and consider $\mathbb{Z}[x]/I[x]$, which is an integral domain, since we know by the Corollary to Lemma 4.6.2 that $\mathbb{Z}[x]/I[x] \simeq (\mathbb{Z}/I)[x] \simeq \mathbb{Z}_p[x]$. The image of $f(x) = x^n + a_1x^{n-1} + \cdots + a_n$ in $\mathbb{Z}[x]/I[x]$ is x^n, since $p \mid a_1, \ldots, p \mid a_n$. So if $\overline{u}(x)$ is the image of $u(x)$ and $\overline{v}(x)$

that of $v(x)$ in $\mathbb{Z}[x]/I[x]$, then $x^n = \bar{u}(x)\bar{v}(x)$. Since $\bar{u}(x)|x^n$, $\bar{v}(x)|x^n$ in $\mathbb{Z}[x]/I[x]$, we must have that $\bar{u}(x) = x^r, \bar{v}(x) = x^{n-r}$ for some $1 \leq r < n$. But then $u(x) = x^r + pg(x)$ and $v(x) = x^{n-r} + ph(x)$, where $g(x)$ and $h(x)$ are polynomials with integer coefficients. Since $u(x)v(x) = x^n + px^r h(x) + px^{n-r}g(x) + p^2 g(x)h(x)$, and since $1 \leq r < n$, the constant term of $u(x)v(x)$ is $p^2 st$, where s is the constant term of $g(x)$ and t the constant term of $h(x)$. Because $f(x) = u(x)v(x)$, their constant terms are equal, hence $a_n = p^2 st$. Since s and t are integers, we get that $p^2 \mid a_n$, a contradiction. In this way we see that $f(x)$ is irreducible. □

We give some examples of the use to which the Eisenstein Criterion can be put.

1. Let $f(x) = x^n - p$, p any prime. Then one sees at a glance that $f(x)$ is irreducible in $\mathbb{Q}[x]$, for the Eisenstein Criterion applies.
2. Let $f(x) = x^5 - 4x + 22$. Since $2 \mid 22$, $2^2 \nmid 22$ and 2 divides the other relevant coefficients of $f(x)$, the Eisenstein Criterion tells us that $f(x)$ is irreducible in $\mathbb{Q}[x]$.
3. Let $f(x) = x^{11} - 6x^4 + 12x^3 + 36x - 6$. We see that $f(x)$ is irreducible in $\mathbb{Q}[x]$ by using either 2 or 3 to check the conditions of the Eisenstein Criterion.
4. Let $f(x) = 5x^4 - 7x + 7$; $f(x)$ is not monic, but we can modify $f(x)$ slightly to be in a position where we can apply the Eisenstein Criterion. Let

   $$g(x) = 5^3 f(x) = 5^4 x^4 - 7 \cdot 5^3 x + 7 \cdot 5^3 = (5x)^4 - 175(5x) + 875.$$

   If we let $y = 5x$, then $g(x) = h(y) = y^4 - 175y + 875$. The polynomial $h(y)$ is irreducible in $\mathbb{Z}[y]$ by using the prime 7 and applying the Eisenstein Criterion. The irreducibility of $h(y)$ implies that of $g(x)$, and so that of $f(x)$, in $\mathbb{Q}[x]$.

   This suggests a slight generalization of the Eisenstein Criterion to nonmonic polynomials. (See Problem 4.)

5. Let $f(x) = x^4 + x^3 + x^2 + x + 1$; as it stands we cannot, of course, apply the Eisenstein Criterion to $f(x)$. We pass to a polynomial $g(x)$ closely related to $f(x)$ whose irreducibility in $\mathbb{Q}[x]$ will ensure that of $f(x)$. Let $g(x) = f(x + 1) = (x + 1)^4 + (x + 1)^3 + (x + 1)^2 + (x + 1) + 1 = x^4 + 5x^3 + 10x^2 + 10x + 5$. The Eisenstein Criterion applies to $g(x)$, using the prime 5; thus $g(x)$ is irreducible in $\mathbb{Q}[x]$. This implies that $f(x)$ is irreducible in $\mathbb{Q}[x]$. (See Problem 1.)

Gotthold Eisenstein (1823–1852) in his short life made fundamental contributions in algebra and analysis.

PROBLEMS

1. In Example 5, show that because $g(x)$ is irreducible in $\mathbb{Q}[x]$, then so is $f(x)$.
2. Prove that $f(x) = x^3 + 3x + 2$ is irreducible in $\mathbb{Q}[x]$.
3. Show that there is an infinite number of integers a such that $f(x) = x^7 + 15x^2 - 30x + a$ is irreducible in $\mathbb{Q}[x]$. What a's do you suggest?
4. Prove the following generalization of the Eisenstein Criterion. Let $f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$ have integer coefficients and suppose that there is a prime p such that

$$p \nmid a_0,\ p \mid a_1,\ p \mid a_2, \ldots, p \mid a_{n-1},\ p \mid a_n,$$

but $p^2 \nmid a_n$; then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

5. If p is a prime, show that $f(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible in $\mathbb{Q}[x]$.
6. Let F be the field and φ an automorphism of $F[x]$ such that $\varphi(a) = a$ for every $a \in F$. If $f(x) \in F[x]$, prove that $f(x)$ is irreducible in $F[x]$ if and only if $g(x) = \varphi(f(x))$ is.
7. Let F be a field. Define the mapping

$$\varphi : F[x] \to F[x] \quad \text{by} \quad \varphi(f(x)) = f(x + 1)$$

for every $f(x) \in F[x]$. Prove that φ is an automorphism of $F[x]$ such that $\varphi(a) = a$ for every $a \in F$.

8. Let F be a field and $b \neq 0$ an element of F. Define the mapping $\varphi : F[x] \to F[x]$ by $\varphi(f(x)) = f(bx)$ for every $f(x) \in F[x]$. Prove that φ is an automorphism of $F[x]$ such that $\varphi(a) = a$ for every $a \in F$.
9. Let F be a field, $b \neq 0$, c elements of F. Define the mapping $\varphi : F[x] \to F[x]$ by $\varphi(f(x)) = f(bx + c)$ for every $f(x) \in F[x]$. Prove that φ is an automorphism of $F[x]$ such that $\varphi(a) = a$ for every $a \in F$.
10. Let φ be an automorphism of $F[x]$, where F is a field, such that $\varphi(a) = a$ for every $a \in F$. Prove that if $f(x) \in F[x]$, then $\deg \varphi(f(x)) = \deg f(x)$.
11. Let φ be an automorphism of $F[x]$, where F is a field, such that $\varphi(a) = a$ for every $a \in F$. Prove there exist $b \neq 0$, c in F such that $\varphi(f(x)) = f(bx + c)$ for every $f(x) \in F[x]$.

12. Find a nonidentity automorphism φ of $\mathbb{Q}[x]$ such that φ^2 is the identity automorphism of $\mathbb{Q}[x]$.

13. Show that in Problem 12 you do not need the assumption $\varphi(a) = a$ for every $a \in \mathbb{Q}$ because any automorphism of $\mathbb{Q}[x]$ automatically satisfies $\varphi(a) = a$ for every $a \in \mathbb{Q}$.

14. Let $\mathbb{C}$ be the field of complex numbers. Given an integer $n > 0$, exhibit an automorphism φ of $\mathbb{C}[x]$ of order n.

7. FIELD OF QUOTIENTS OF AN INTEGRAL DOMAIN

Given the integral domain $\mathbb{Z}$, the ring of integers, then intimately related to $\mathbb{Z}$ is the field $\mathbb{Q}$ of rational numbers that consists of all fractions of integers; that is, all quotients m/n, where $m, n \neq 0$ are in $\mathbb{Z}$. Note that there is no unique way of representing $\frac{1}{2}$, say, in $\mathbb{Q}$ because $\frac{1}{2} = \frac{2}{4} = (-7)/(-14) = \cdots$. In other words, we are identifying $\frac{1}{2}$ with $\frac{2}{4}$, $(-7)/(-14)$, and so on. This suggests that what is really going on in constructing the rationals from the integers is some equivalence relation on some set based on the integers.

The relation of $\mathbb{Q}$ to $\mathbb{Z}$ carries over to any integral domain D. Given an integral domain D, we shall construct a field $F \supset D$ whose elements will be quotients a/b with $a, b \in D$, $b \neq 0$. We go through this construction formally.

Let D be an integral domain, and let $S = \{(a, b) \mid a, b \in D, b \neq 0\}$; S is thus the subset of $D \times D$—the Cartesian product of D with itself—in which the second component is not allowed to be 0. Think of (a, b) as a/b for a moment; if so, when would we want to declare that $(a, b) = (c, d)$? Clearly, we would want this if $a/b = c/d$, which in D itself would become $ad = bc$. With this as our motivating guide we define a relation $\sim$ on S by declaring:

$$(a, b) \sim (c, d) \text{ for } (a, b), (c, d) \text{ in } S, \text{ if and only if } ad = bc.$$

We first assert that this defines an equivalence relation on S. We go through the three requirements for an equivalence relation term by term.

1. $(a, b) \sim (a, b)$, for clearly $ab = ba$ (since D is commutative). So $\sim$ is reflexive.
2. $(a, b) \sim (c, d)$ implies that $(c, d) \sim (a, b)$, for $(a, b) \sim (c, d)$ means $ad = bc$; for $(c, d) \sim (a, b)$ we need $cb = da$, but this is true, since $cb = bc = ad = da$. So $\sim$ is symmetric.
3. $(a, b) \sim (c, d)$, $(c, d) \sim (e, f)$ implies that $ad = bc$, $cf = de$, so $adf = bcf = bde$; but $d \neq 0$ and we are in an integral domain, hence $af = be$ follows. This says that $(a, b) \sim (e, f)$. So the relation is transitive.

We have shown that $\sim$ defines an equivalence relation on S. Let F be the set of all the equivalence classes $[a, b]$ of the elements $(a, b) \in S$. F is our required field.

To show that F is a field, we must endow it with an addition and multiplication. First the addition; what should it be? Forgetting all the fancy talk about equivalence relation and the like, we really want $[a, b]$ to be a/b. If so, what should $[a, b] + [c, d]$ be other than the formally calculated

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}?$$

This motivates us to *define*

$$[a, b] + [c, d] = [ad + bc, bd]. \tag{1}$$

Note that since $b \neq 0$, $d \neq 0$ and D is a domain, then $bd \neq 0$, hence $[ad + bc, bd]$ is a legitimate element of F.

As usual we are plagued with having to show that the addition so defined in F is well-defined. In other words, we must show that if $[a, b] = [a', b']$ and $[c, d] = [c', d']$, then $[a, b] + [c, d] = [a', b'] + [c', d']$. From (1) we must thus show that $[ad + bc, bd] = [a'd' + b'c', b'd']$, which is to say, $(ad + bc)b'd' = bd(a'd' + b'c')$. Since $[a, b] = [a', b']$ and $[c, d] = [c', d']$, $ab' = ba'$ and $cd' = dc'$. Therefore, $(ad + bc)b'd' = ab'dd' + bb'cd' = ba'dd' + bb'dc' = (a'd' + b'c')bd$, as required. Thus "+" is well-defined in F.

The class $[0, b]$, $b \neq 0$, acts as the 0 under "+," we denote it simply as 0, and the class $[-a, b]$ is the negative of $[a, b]$. To see that this makes of F an abelian group is easy, but laborious, for all that really needs verification is the associative law.

Now to the multiplication in F. Again motivated by thinking of $[a, b]$ as a/b, we *define*

$$[a, b][c, d] = [ac, bd]. \tag{2}$$

Again since $b \neq 0$, $d \neq 0$, we have $bd \neq 0$, so the element $[ac, bd]$ is also a legitimate element of F.

As for the "+" we must show that the product so introduced is well-defined; that is, if $[a, b] = [a', b']$, $[c, d] = [c', d']$, then

$$[a, b][c, d] = [ac, bd] = [a'c', b'd'] = [a', b'][c', d'].$$

We know that $ab' = ba'$ and $cd' = dc'$, so $acb'd' = ab'cd' = ba'dc' = bda'c'$, which is exactly what we need for $[ac, bd] = [a'c', b'd']$. Thus the product is well-defined in F.

What acts as 1 in F? We claim that for any $a \neq 0$, $b \neq 0$ in D, $[a, a] = [b, b]$ (since $ab = ba$) and $[c, d][a, a] = [ca, da] = [c, d]$, since $(ca)d = (da)c$.

So $[a, a]$ acts as 1, and we write it simply as $1 = [a, a]$ (for all $a \neq 0$ in D). Given $[a, b] \neq 0$, then $a \neq 0$, so $[b, a]$ is in F; hence, because $[a, b][b, a] = [ab, ba] = [ab, ab] = 1$, $[a, b]$ has an inverse in F. All that remains to show that the nonzero elements of F form an abelian group under this product is the associative law and commutative law. We leave these to the reader.

To clinch that F is a field, we need only now show the distributive law. But $[ad + bc, bd][e, f] = [(ad + bc)e, bdf]$, so

$$([a, b] + [c, d])[e, f] = [ade + bce, bdf],$$

while

$$\begin{aligned}
&[a, b][e, f] + [c, d][e, f] \\
&\quad = [ae, bf] + [ce, df] = [aedf + bfce, bdf^2] \\
&\quad = [(ade + bce) f, bdf^2] = [ade + bce, bdf][f, f] \\
&\quad = [ade + bce, bdf],
\end{aligned}$$

which we have seen is $([a, b] + [c, d])[e, f]$. The distributive law is now established, so F is a field.

Let $a \neq 0$ be a fixed element in D and consider $[da, a]$ for any $d \in D$. The map $\varphi : d \to [da, a]$ is a monomorphism of D into F. It is certainly 1-1, for if $\varphi(d) = [da, a] = 0$, then $da = 0$, so $d = 0$, since D is an integral domain. Also, $\varphi(d_1 d_2) = [d_1 d_2 a, a]$ while $\varphi(d_1)\varphi(d_2) = [d_1 a, a][d_2 a, a] = [d_1 d_2 a^2, a^2] = [d_1 d_2 a, a][a, a] = [d_1 d_2 a, a] = \varphi(d_1 d_2)$. Furthermore,

$$\begin{aligned}
[d_1 a, a] + [d_2 a, a] &= [d_1 a^2 + a^2 d_2, a^2] \\
&= [d_1 a + d_2 a, a][a, a] \\
&= [(d_1 + d_2) a, a]
\end{aligned}$$

so $\varphi(d_1 + d_2) = [(d_1 + d_1)a, a] = [d_1 a, a] + [d_2 a, a] = \varphi(d_1) + \varphi(d_2)$. Thus φ maps D monomorphically into F. So, D is isomorphic to a subring of F, and we can thus consider D as "embedded" in F. We consider every element $[a, b]$ of F as the fraction a/b.

Theorem 4.7.1. Let D be an integral domain. Then there exists a field $F \supset D$ which consists of all fractions a/b, as defined above, of elements in D.

The field F is called the *field of quotients* of D. When $D = \mathbb{Z}$, then F is isomorphic to the field $\mathbb{Q}$ of rational numbers. Also, if D is the domain of even integers, then F is also the entire field $\mathbb{Q}$.

What we did above in constructing the field of quotients of D was a long, formal, wordy, and probably dull way of doing something that is by its

nature something very simple. We really are doing nothing more than forming all formal fractions a/b, $a, b \neq 0$ in D, where we add and multiply fractions as usual. However, it is sometimes necessary to see something done to its last detail, painful though it may be. Most of us had never seen a really formal and precise construction of the rationals from the integers. Now that we have constructed F from D in this formal manner, forget the details and think of F as the set of all fractions of elements of D.

PROBLEMS

1. Prove the associative law of addition in F.

2. Prove the commutative law of addition in F.

3. Prove that the product in F is commutative and associative.

4. If K is any field that contains D, show that $K \supset F$. (So F is the smallest field that contains D.)